VOLUME 102 PART C NUMBER 2 SEPTEMBER 1955

The Proceedings

OF

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART C

MONOGRAPHS Nos. 117–131

SAVOY PLACE · LONDON W.C.2

Price Eight Shillings and Sixpence

The Institution of Electrical Engineers

FOUNDED 1871
INCORPORATED BY ROYAL CHARTER 1921

PATRON: HER MAJESTY THE QUEEN

COUNCIL 1954–1955

President

J. ECCLES, C.B.E., B.Sc.

Past-Presidents

SIR JAMES SWINBURNE, BART., F.R.S.
W. H. ECCLES, D.Sc., F.R.S.
THE RT. HON. THE EARL OF MOUNT EDGCUMBE, T.D.
J. M. DONALDSON, M.C.
PROFESSOR E. W. MARCHANT, D.Sc.
P. V. HUNTER, C.B.E.
H. T. YOUNG.
SIR GEORGE LEE, O.B.E., M.C.
SIR ARTHUR P. M. FLEMING, C.B.E., D.Eng., LL.D.
J. R. BEARD, C.B.E., M.Sc.
SIR NOEL ASHBRIDGE, B.Sc.(Eng.).
COLONEL SIR A. STANLEY ANGWIN, K.B.E., D.S.O., M.C., T.D., D.Sc.(Eng.).
SIR HARRY RAILING, D.Eng.
P. DUNSHEATH, C.B.E., M.A., D.Sc.(Eng.).
SIR VINCENT Z. DE FERRANTI, M.C.
T. G. N. HALDANE, M.A.
PROFESSOR E. B. MOULLIN, M.A., Sc.D.
SIR ARCHIBALD J. GILL, B.Sc.(Eng.).
SIR JOHN HACKING.
COLONEL B. H. LEESON, C.B.E., T.D.
SIR HAROLD BISHOP, C.B.E., B.Sc.(Eng.).

Vice-Presidents

T. E. GOLDUP, C.B.E.
S. E. GOODALL, M.Sc.(Eng.).
WILLIS JACKSON, D.Sc., D.Phil., F.R.S.
SIR GEORGE H. NELSON, BART.
SIR W. GORDON RADLEY, C.B.E., Ph.D.(Eng.).

Honorary Treasurer

H. W. GRIMMITT.

Ordinary Members of Council

J. BENNETT.
A. R. COOPER.
A. T. CRAWFORD, B.Sc.
C. DANNATT, O.B.E., D.Sc.
B. DONKIN, B.A.
O. W. HUMPHREYS, B.Sc.
C. R. KING, C.B.E.
H. R. L. LAMONT, Ph.D., M.A., B.Sc.
F. J. LANE, O.B.E., M.Sc.
G. S. C. LUCAS, O.B.E.
G. LYON, M.Sc.(Eng.).
SIR HAMISH D. MACLAREN, K.B.E., C.B., D.F.C., LL.D., B.Sc.
A. H. MUMFORD, O.B.E., B.Sc.(Eng.).
W. F. PARKER.
PROFESSOR M. G. SAY, Ph.D., M.Sc.
R. L. SMITH-ROSE, C.B.E., D.Sc., Ph.D.
G. O. WATSON.
J. H. WESTCOTT, B.Sc.(Eng.), Ph.D.
E. L. E. WHEATCROFT, M.A.
R. T. B. WYNN, C.B.E., M.A.

Chairmen and Past-Chairmen of Sections

Measurements:
M. WHITEHEAD.
*J. F. COALES, O.B.E., M.A.

Radio:
C. W. OATLEY, M.A., M.Sc.
*J. A. SMALE, C.B.E., A.F.C., B.Sc.

Supply:
J. D. PEATTIE, B.Sc.
*L. G. BRAZIER, Ph.D., B.Sc.

Utilization:
J. I. BERNARD, B.Sc.Tech.
*B. L. METCALF, B.Sc.(Eng.).

Chairmen and Past-Chairmen of Local Centres

East Midland Centre:
J. M. MITCHELL, B.Sc., Ph.D.
*C. D. WILKINSON.

Mersey and North Wales Centre:
P. R. DUNN, B.Sc.
*T. COATES, M.Eng.

North Midland Centre:
W. A. CROCKER.
*G. CATON.

North-Eastern Centre:
G. W. B. MITCHELL, B.A.
*H. ESTHER, B.Eng.

North-Western Centre:
PROFESSOR E. BRADSHAW, M.B.E., M.Sc.Tech., Ph.D.
*H. WEST.

Northern Ireland Centre:
MAJOR P. L. BARKER, B.Sc.
*J. R. W. MURLAND, B.Sc.(Eng.).

Western Centre:
A. N. IRENS.
*J. VAUGHAN HARRIES.

Scottish Centre:
J. S. HASTIE, B.Sc.(Eng.).
*C. H. A. COLLYNS.

South Midland Centre:
A. R. BLANDFORD.
*H. J. GIBSON, B.Sc.

Southern Centre:
E. A. LOGAN, M.Sc.
*COMDR. C. V. ROBINSON, R.N., O.B.E.

* Past-Chairman.

Secretary

W. K. BRASHER, C.B.E., M.A., M.I.E.E.

Deputy Secretary
F. JERVIS SMITH, M.I.E.E.

Assistant Secretary
F. C. HARRIS.

Editor-in-Chief
G. E. WILLIAMS, B.Sc.(Eng.), M.I.E.E.

THE PROCEEDINGS OF
THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

VOL. 102. PART C. NO. 2. SEPTEMBER 1955

621.313.1.011

Monograph No. 117 S
Jan. 1955

THE TENSOR EQUATIONS OF ELECTRICAL MACHINES

By J. W. LYNN, M.Sc., Associate Member.

(*The paper was first received 28th December, 1953, and in revised form 18th August, 1954. It was published as an* INSTITUTION MONOGRAPH *in January, 1955.*)

SUMMARY

In the analysis of modern electrical machine systems there has been a trend towards the use of various components of the variables and parameters of the system, examples being the use of symmetrical components and direct- and quadrature-axis quantities. Such quantities may be considered to be different "reference frames" to which the various currents, fluxes, etc., are referred. Kron has shown that such transformations may be expressed in a general way when the equations of a network or machine are written in tensor form.

Transformations may then be carried out to any one of a number of reference axes, which may be stationary or rotating with respect to the windings of the machine or network. In particular it has been found that in certain cases there are advantages in analysing machines with respect to axes rotating with the flux.

The purpose of the present paper is to investigate the tensor form of the equations of electrical machines, to demonstrate the differences between tensor and non-tensor terms and to show how these terms are interpreted in application to simple cases. The equations of a primitive machine are examined in both stationary and rotating axes and the equations of a 3-phase series impedance and a 3-phase induction-motor are derived in both systems of reference.

LIST OF PRINCIPAL SYMBOLS

Tensor Notation

Indices:

a, b, c = Quantities in axes fixed to the machine stator and rotor windings.

k, n, m = Quantities in axes all relatively stationary.

α, β, γ = Quantities in axes fixed or free on the stator and rotating freely on the rotor.

s, t = Quantities associated with the mechanical rotational effects in the machine (e.g. generated voltages and torque).

u, v, w = Quantities in a general equation.

Electrical parts of the equations:

f_m, f_α, etc. = Electrical voltage vectors in axes denoted by indices.

x^k, x^α, etc. = Electric variables. The electrical charges in machine windings, referred to axes denoted by indices.

$\dot{x}^\alpha$ (equivalent to i^α) = Electric current vector, in axes denoted by indices.

$R_{\gamma\alpha}$ = Resistance matrix, in axes denoted by indices.

$L_{\gamma\alpha}$ = Inductance matrix, in axes denoted by indices.

$G_{\gamma\alpha}$ = Generated voltage coefficients, in axes denoted by indices.

Mechanical part of the equations:

f_s = Mechanical force.

x^s = Mechanical variable θ, the angular position of the machine rotor during rotation.

$\dot{x}^s$ (equivalent to i^s) = Angular velocity of machine rotor.

R_{st} = Mechanical friction coefficients.

L_{st} = Moment of inertia of machine rotor.

General symbols:

$\boldsymbol{e}$ = Generalized force vector (voltage or mechanical force).

$\boldsymbol{R}$ = Generalized dissipation matrix (resistance or friction).

$\boldsymbol{L}$ = Generalized inductance matrix (inductance or inertia).

$\boldsymbol{i}$ = Generalized current vector (electric current or angular velocity).

C_a^k = Connection matrix between quantities in axes denoted by indices.

$\boldsymbol{C}$ = Direct notation for C_a^k, etc.

$\boldsymbol{C}_t$ = Transpose of matrix $\boldsymbol{C}$.

$\Omega_{\alpha\beta}^{\delta}$ = "Non-holonomic object" containing functions of C, in axes denoted by indices.

$[\alpha\beta, \gamma]$ = A "connection" term containing functions of $L_{\gamma\alpha}$ in axes denoted by indices.

$\Gamma_{\alpha\beta,\gamma}$ = A "connection" term containing both $[\alpha\beta, \gamma]$ and $\Omega_{\alpha\beta,\gamma}$ in axes denoted by indices.

T_{knm} = Rotational or torsion tensor containing the anti-symmetrical part of $\Gamma_{kn,m}$ in axes denoted by indices.

S_{knm} = Tensor components of T_{knm} giving the terms G_{mk}.

$\boldsymbol{B}$ = Flux-density matrix, a tensor.

$\boldsymbol{\varphi}$ = Flux-density matrix, a non-tensor.

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Lynn is in the Electrical Engineering Department, University of Liverpool.

Synchronous Machine Notation

e_f = Field voltage.
e_d = Direct-axis terminal voltage.
e_q = Quadrature-axis terminal voltage.
i^d = Direct-axis current.
i^q = Quadrature-axis current.
R_f = Field resistance.
R_d = Armature resistance in direct axis.
R_q = Armature resistance in quadrature axis.
L_f = Self-inductance of field winding.
L_d = Armature self-inductance in direct axis.
L_q = Armature self-inductance in quadrature axis.
M = Mutual inductance in direct axis (field-armature).

(1) INTRODUCTION

The recent increase in complexity of electrical power networks such as control systems and interconnected power-transmission systems has led to the introduction of systematic methods of analysing the behaviour of networks and machines. The powerful methods of symmetrical components brought about rapid methods of solution of problems associated with unbalanced polyphase circuits; the introduction of the two-reaction theory into synchronous-machine studies has simplified many aspects of the analysis of power-transmission systems. In dealing with complicated problems the trend has been towards the use of components of the quantities involved which, while they are entirely fictitious, lead to elegant solutions.[17,18] In such cases the actual variables and parameters of the system (currents, impedances, etc.) are transformed into the required components.

In 1934 Kron, in America, developed a technique[1] for dealing systematically with such transformations. In his analysis Kron ensures that the electrical power in a network is invariant under a given transformation to new components. He does this by introducing the methods of the tensor calculus. Tensors are sets of quantities which are functions of a set of variables and are subject to certain laws of transformation when the variables are changed (see Appendix 10.1). The subject was developed in the field of the geometry of generalized spaces and consequently the terminology in the literature is largely geometrical; for example, the variables are referred to as co-ordinates. The following two properties of tensors are invaluable in application to any group of transformations:

(*a*) A set of functions usually written in the form g_{ab} is associated with tensor transformations, where a and b range over the n variables. This set of functions determines an invariant, which has the same magnitude in all reference frames.

The element of the invariant is usually written in the literature as dl and then

$$(dl)^2 = g_{ab}dx^a dx^b$$

where dx^a and dx^b are differentials of the variables. The use of this tensor is demonstrated in Appendix 10.1.

(*b*) If an entity is a tensor and exists with a non-zero value in one reference frame (or system of variables) then it has non-zero values, usually with different components, in all reference frames. Quantities which are not tensors may arise in one system of reference and become zero in another; in other words, they may arise because of the particular reference frame chosen. This effect occurs especially when there is relative motion between the reference axes.

Kron takes a primitive or elementary representative network which has comparatively simple equations; these are written in tensor form and transformed to give the equations of the required system.

A wide range of machines can be considered as a group of interconnections of the windings of the primitive machine shown in Fig. 2, and the equations of any of these may be obtained by transformation of the primitive-machine equations.

The expression "reference frame" denotes the system of measurement from which the variables and parameters are determined. In a static electrical network a change of components from branch to mesh currents is a change of reference frame.[2] In the case of a slip-ring induction motor, stator currents will be measured from stationary terminals and rotor currents from axes rotating with the rotor. In commutator machines the reference axes are relatively stationary. Park's application of Blondel's two-reaction theory simplified synchronous-machine theory by converting the stator and rotor quantities to relatively stationary axes. Kron's analysis deals systematically with equations of machine systems having either stationary or relatively moving reference axes, the transformations in all cases following the routine laws of matrix algebra and tensor calculus. His work has introduced to engineers wider concepts of transformation, invariance and theory of groups, all of which have been invaluable in analysis of complex systems in various branches of physics. The purpose of the present paper is twofold, namely to bring together some of the scattered works dealing with Kron's methods and present a continuous account of the development of the subject, and to analyse in detail the distinction between tensor and non-tensor terms in machine equations.

The analysis starts from the dynamical equation of Lagrange.[5] As is shown in Appendix 10.2, Lagrange's equation gives the relation between the potential and kinetic energies in a system and the applied forces, in terms of generalized co-ordinates. In electrical form the corresponding relations are those between the magnetic energy and the applied voltages, in terms of generalized variables which are the electrical charges in the network. This equation is very suitable for certain classes of transformations of co-ordinates, but it has been found that under the conditions of transformation obtaining in electrical machines a modified form of Lagrange's equation must be used. The modified equation was developed by Boltzmann and Hamel[13] to cover certain conditions of constraint in dynamical systems and, as Kron has shown,[6] the Boltzmann–Hamel equation can be used to form a basis for tensor analysis of electrical machines from the dynamical viewpoint.

(2) MACHINE EQUATIONS

The first type of primitive electrical machine to be considered is shown in Fig. 1. The rotor is assumed to be smooth and to

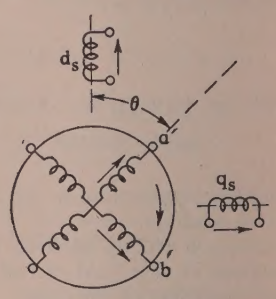


Fig. 1.—Primitive machine with axes fixed to windings.

have on it a symmetrical 2-phase winding sinusoidally distributed. The field is fixed in space and consists of windings d and qs in the stator—direct and quadrature axes respectively. Iron loss and saturation are neglected. The armature axes a′ and b′ are fixed to the armature and rotate with it. Three-phase machines may be analysed by resolving the resultant armature current and flux vectors along two similar axes.[7]

The inductance of phase a′ of this machine may be written:

$$\text{Phase a' self-inductance} = L_A + L_B \cos 2\theta$$

where $L_A = (L_{dr} + L_{qr})/2$

$L_B = (L_{dr} - L_{qr})/2$

L_{dr} and L_{qr} are the self-inductances of rotor phase a′ when in the direct- and quadrature-axis positions respectively. The corresponding values of mutual inductance, rotor to stator, are M_d and M_q.

The equations of this machine may be derived from the dynamical equation of Lagrange:[4]

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{x}^c}\right) - \frac{\partial T}{\partial x^c} + \frac{\partial F}{\partial \dot{x}^c} = f_c \quad . \quad . \quad . \quad . \quad . \quad (2)$$

In electrical machines the generalized variables x^c are the electrical charges in the circuits and the rotor angle θ. The quantities $\dot{x}^c$ therefore represent the currents i^c and the rotor angular velocity $d\theta/dt$. T, the stored magnetic energy, is given by

$$T = \tfrac{1}{2}L_{ab}\frac{dx^a}{dt}\frac{dx^b}{dt} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

F, the dissipative force, is given by

$$F = \tfrac{1}{2}R_{ab}\frac{dx^a}{dt}\frac{dx^b}{dt} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

For the above machine[4]

$$\begin{aligned} T = {} & \tfrac{1}{2}(L_A + L_B\cos 2\theta)(i^{a'})^2 + \tfrac{1}{2}(L_A - L_B\cos 2\theta)(i^{b'})^2 \\ & + \tfrac{1}{2}L_{ds}(i^{ds})^2 - L_B\sin 2\theta i^{a'}i^{b'} + M_d\cos\theta i^{ds}i^{a'} \\ & - M_d\sin\theta i^{ds}i^{b'} + M_q\sin\theta i^{qs}i^{b'} + M_q\cos\theta i^{qs}i^{b'} \\ & + \tfrac{1}{2}L_{qs}(i^{qs})^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5) \end{aligned}$$

$$F = \tfrac{1}{2}[R_{a'}(i^{a'})^2 + R_{b'}(i^{b'})^2 + R_{ds}(i^{ds})^2 + R_{qs}(i^{qs})^2] \quad (6)$$

Substituting eqns. (5) and (6) in eqn. (2), the complete set of equations for the machine may be written

$$\boldsymbol{e} = \boldsymbol{Ri} + p\boldsymbol{Li} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where $\left(p \equiv \frac{d}{dt}\right)$

In matrix form these are[2]

			ds	a′	b′	qs		
ds	e_{ds}	ds	$R_{ds} + pL_{ds}$	$pM_d\cos\theta$	$-pM_d\sin\theta$		ds	i^{ds}
a′	e_a	a′	$pM_d\cos\theta$	$R_{dr} + p(L_{dr}\cos^2\theta + L_{qr}\sin^2\theta)$	$p(L_{qr} - L_{dr})\sin\theta\cos\theta$	$pM_q\sin\theta$	a′	$i^{a'}$
b′	e_b	b′	$-pM_d\sin\theta$	$p(L_{qr} - L_{dr})\sin\theta\cos\theta$	$R_{qr} + p(L_{dr}\sin^2\theta + L_{qr}\cos^2\theta)$	$pM_q\cos\theta$	b′	$i^{b'}$
qs	e_{qs}	qs		$pM_q\sin\theta$	$pM_q\cos\theta$	$R_{qs} + pL_{qs}$	qs	i^{qs}

. . . . (8)

Note.—The index associated with each current value is written as a superscript since this is required by the index notation used in tensor calculus.

It is shown in Appendix 10.3 that in tensor form the equations of Lagrange are written

$$L_{cb}\frac{d^2x^b}{dt^2} + [ab,c]\frac{dx^a}{dt}\frac{dx^b}{dt} + R_{cb}\frac{dx^b}{dt} = f_c \quad . \quad . \quad (10)$$

where $$[ab,c] \equiv \frac{1}{2}\left(\frac{\partial L_{cb}}{\partial x^a} + \frac{\partial L_{ca}}{\partial x^b} - \frac{\partial L_{ab}}{\partial x^c}\right) \quad . \quad . \quad . \quad (11)$$

The voltage equation is obtained by allowing the free index c to cover the electrical range of variables, the other indices ranging over the electrical and mechanical parts:[1]

$$e_c = L_{cb}\frac{d^2x^b}{dt^2} + [as,c]\frac{dx^a}{dt}\frac{dx^s}{dt} + [sb,c]\frac{dx^s}{dt}\frac{dx^b}{dt} + R_{cb}\frac{dx^b}{dt} \quad . \quad (12)$$

or $$e_c = L_{cb}\frac{di^b}{dt} + [as,c]i^ai^s + [sb,c]i^si^b + R_{cb}i^b \quad . \quad . \quad (13)$$

From eqn. (11),

$$[as,c] = \frac{1}{2}\left(\frac{\partial L_{cs}}{\partial x^a} + \frac{\partial L_{ca}}{\partial x^s} - \frac{\partial L_{as}}{\partial x^c}\right)$$

Since there is no mutual coupling between an electrical row and a mechanical column

$$\frac{\partial L_{cs}}{\partial x^a} = \frac{\partial L_{as}}{\partial x^c} = 0$$

$$[as,c]i^ai^s = \frac{1}{2}\frac{\partial L_{ca}}{\partial x^s}i^a\frac{dx^s}{dt} = \frac{1}{2}\frac{dL_{ca}}{dt}i^a \quad . \quad . \quad . \quad (14)$$

$$[sb,c]i^si^b = \frac{1}{2}\frac{dL_{cb}}{dt}i^b \quad . \quad . \quad . \quad . \quad . \quad (15)$$

Thus Lagrange's equations give

$$e_c = R_{cb}i^b + L_{cb}\frac{di^b}{dt} + \frac{dL_{cb}}{dt}i^b \quad . \quad . \quad . \quad (16)$$

or $$e_c = R_{cb}i^b + \frac{d}{dt}(L_{cb}i^b) \quad . \quad . \quad . \quad . \quad (17)$$

This is, of course, Maxwell's equation as shown in eqn. (7). The equation of torque is obtained by allowing the free index

From eqn. (8) the inductance matrix may be written as below. An additional mechanical row and column s may be added to include the mechanical inertia.

The general inductance matrix is therefore:

	ds	a′	b′	qs	s
ds	L_{ds}	$-M_d\cos\theta$	$-M_d\sin\theta$		
a′	$M_d\cos\theta$	$L_{dr}\cos^2\theta + L_{qr}\sin^2\theta$	$(L_{qr} - L_{dr})\sin\theta\cos\theta$	$M_q\sin\theta$	
b′	$-M_d\sin\theta$	$(L_{qr} - L_{dr})\sin\theta\cos\theta$	$L_{dr}\sin^2\theta + L_{qr}\cos^2\theta$	$M_q\cos\theta$	.
qs		$M_q\sin\theta$	$M_q\cos\theta$	L_{qs}	
s	.				L_{ss}

. . . . (9)

where L_{ss} is the moment of inertia of the rotor.

to cover only the mechanical part of the range of variables,[1]

thus $$f_s = R_{st}i^t + L_{st}\frac{d^2x^t}{dt^2} + [ab,s]i^a i^b \quad . \quad . \quad . \quad (18)$$

where $$[ab,s] = -\frac{1}{2}\frac{\partial L_{ab}}{\partial \theta}$$

Therefore $$f_s = R_{st}\frac{d\theta}{dt} + L_{st}\frac{d^2\theta}{dt^2} - \frac{1}{2}\frac{\partial L_{ab}}{\partial \theta}i^a i^b \quad . \quad . \quad . \quad (19)$$

$R_{st}\frac{d\theta}{dt}$ = frictional torque

$L_{st}\frac{d^2\theta}{dt^2}$ = inertia torque

and $-\frac{1}{2}\frac{\partial L_{ab}}{\partial \theta}i^a i^b$ = electrical torque

The second form of the primitive machine considered here is shown in Fig. 2. The rotating axes a' and b' of Fig. 1 are here

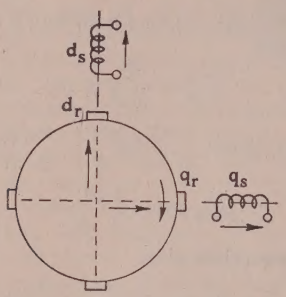


Fig. 2.—Primitive machine with stationary axes.

resolved along the direct and quadrature axes. All axes are now relatively stationary. This primitive machine has been used by Sabbagh, Stanley, Kron and others[7–11] as the basis of the analysis of many derived machines. The equations may be written down by inspection when the resistance voltage-drops and induced and generated voltages are considered. In matrix form these are written

$$\begin{array}{c|c} & \\ ds \\ dr \\ qr \\ qs \end{array}\!\!\begin{array}{|c|}\hline e_{ds} \\ e_{dr} \\ e_{qr} \\ e_{qs} \\ \hline \end{array} = \begin{array}{c|c|c|c|c} & ds & dr & qr & qs \\ ds & R_{ds}+L_{ds}p & M_d p & & \\ dr & M_d p & R_{dr}+L_{dr}p & L'_{qr}p\theta & M'_q p\theta \\ qr & -M'_d p\theta & -L'_{dr}p\theta & R_{qr}+L_{qr}p & M_q p \\ qs & & & M_q p & R_{qs}+L_{qs}p \end{array} \cdot \begin{array}{c|c} & \\ ds & i^{ds} \\ dr & i^{dr} \\ qr & i^{qr} \\ qs & i^{qs} \end{array} \quad . \quad . \quad . \quad . \quad (20)$$

The voltage vector contains impressed voltages in all axes. The generated voltage coefficients are written, for example, $M'_d p\theta$ as compared with $M_d p$ for the induced voltages since the flux waveform may not be sinusoidal. When the flux waveform is sinusoidal the maximum voltage generated by rotation of a coil in this flux at synchronous speed will be equal to the maximum voltage induced by this flux linking the coil and $M'_d p\theta i$ is equal to $M_d pi$ and $M'_d = M_d$. Eqns. (20) are seen to include those for a synchronous machine according to Park's two-reaction theory.[18] These as used by Concordia[12] are written:

Impressed field voltage, $e_f = i^f R_f + L_f pi^f + M_d pi^{dr}$

Generated voltage,
$$e_d = -M_d pi^f - R_d i^{dr} - L_d pi^{dr} + L_q p\theta i^{qr}$$

Generated voltage,
$$e_q = -M_d p\theta i^f - L_d p\theta i^{dr} - R_q i^{qr} - L_q pi^{qr} \quad . \quad (21)$$

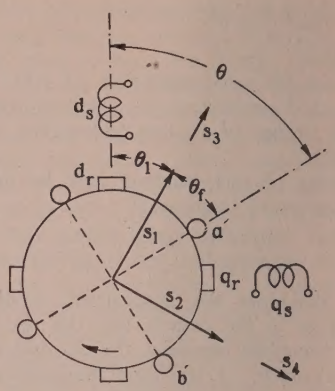


Fig. 3.—Primitive machine with axes rotating freely.

In synchronous-machine studies the quadrature-axis stator (field) coils are omitted unless amortisseur windings are being considered. Here sinusoidal flux distribution is considered and $M'_d = M_d$, etc.

Eqns. (20) may be obtained from those of the previous form of primitive machine using the relationship

$$\left.\begin{aligned} i^{a'} &= i^{dr}\cos\theta + i^{qr}\sin\theta \\ i^{b'} &= i^{qr}\cos\theta - i^{dr}\sin\theta \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (22)$$

In index notation the stationary-axis equations may be written[2]

$$\text{Voltage, } e_m = R_{mn}i^n + L_{mn}\frac{di^n}{dt} + G_{mn}i^n p\theta \quad . \quad . \quad (23)$$

$$\text{Torque, } f_s = R_{st}i^t + L_{st}\frac{d^2\theta}{dt^2} - G_{mn}i^m i^n \quad . \quad . \quad (24)$$

Using the concepts of the tensor calculus these two equations are written by Kron[2] as one electro-mechanical equation,

$$e_w = R_{wv}i^v + L_{wv}\frac{di^v}{dt} + \Gamma_{uv,w}i^u i^v \quad . \quad . \quad . \quad (25)$$

$$\text{voltage } e_m = R_{mn}i^n + L_{mn}\frac{di^n}{dt} + \Gamma_{ks,m}i^k i^s + \Gamma_{sn,m}i^s i^n \quad . \quad (26)$$

$$\text{torque } f_s = R_{st}i^t + L_{st}\frac{di^t}{dt} + \Gamma_{kn,s}i^k i^n \quad . \quad . \quad (27)$$

The equation of Maxwell, eqn. (17), does not give these equations directly since it does not include generated voltages. The following Sections show that the connection $\Gamma_{uv,w}$ arises naturally because of the dynamical relationship between the two types of primitive machine, this being quasi-holonomic and non-integrable. Fig. 4 shows the form of the connection $\Gamma_{uv,w}$ when written as a matrix in the form of a cube, together with the arrangement of matrix multiplication leading from eqns. (26) and (27) to eqns. (23) and (24).

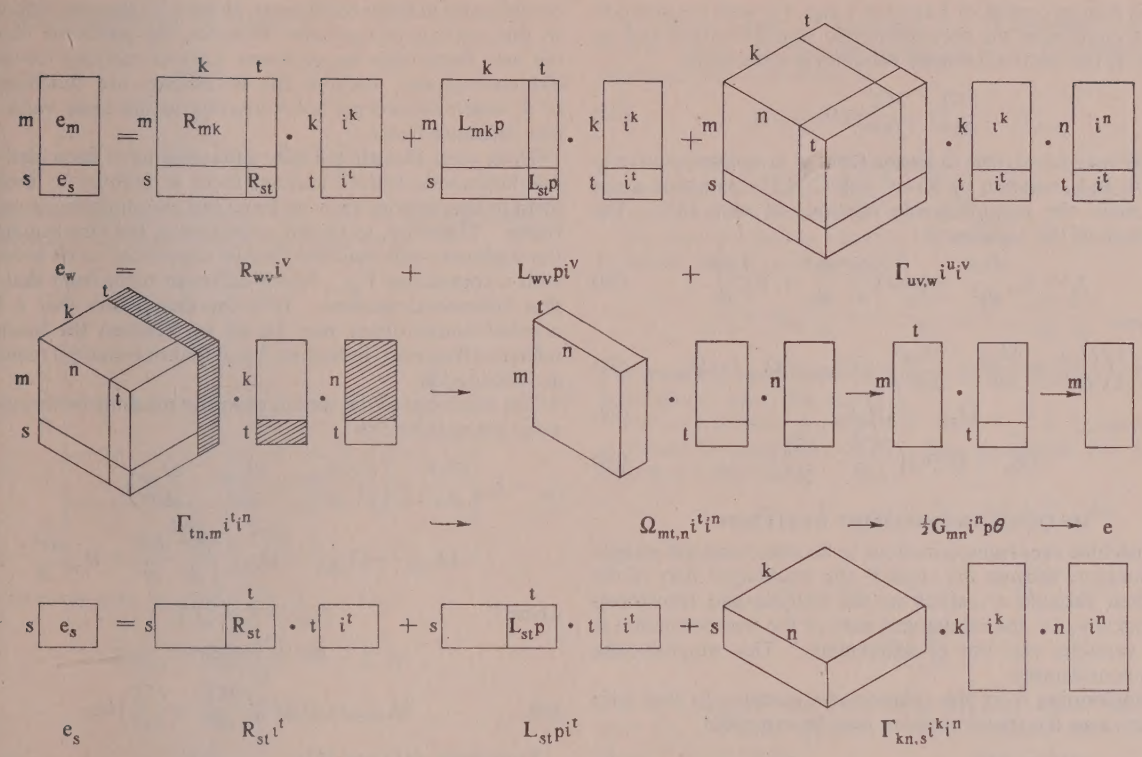


Fig. 4.—Machine tensor equation written in matrix form.

(3) NON-HOLONOMIC TRANSFORMATIONS

The currents in the armature axes a′ and b′ in Fig. 1 may be resolved along "d" and "q" axes shown in Fig. 2, the relationship being

$$\left.\begin{aligned} i^{dr} &= i^{a'}\cos\theta - i^{b'}\sin\theta \\ i^{qr} &= i^{a'}\sin\theta + i^{b'}\cos\theta \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (28)$$

Since the variables in Lagrange's equations are the charges, x^a, eqns. (28) represent a transformation of differentials of the variables, where $dx^a/dt = i^a$, etc. These are equations of constraint among the differentials of the variables x^k, and the transformation must therefore be written

$$\left.\begin{aligned} dx^{dr} &= dx^{a'}\cos\theta - dx^{b'}\sin\theta \\ dx^{qr} &= dx^{a'}\sin\theta + dx^{b'}\cos\theta \end{aligned}\right\} \quad . \quad . \quad . \quad (29)$$

They obtain at a given instant and cannot be integrated to give a relationship among the charges. There is no corresponding relationship

$$x^{dr} = f(x^{a'}, x^{b'}, \theta)$$

If such a function existed the following results would be obtained:[2]

$$dx^{dr} = \frac{\partial f^{dr}}{\partial x^{a'}}dx^{a'} - \frac{\partial f^{dr}}{\partial x^{b'}}dx^{b'}$$

Thus if $$\frac{\partial f^{dr}}{\partial x^{a'}} = C^{dr}_{a'}$$

Then $$\frac{\partial C^{dr}_{a'}}{\partial\theta} = -\sin\theta = \frac{\partial f^{dr}}{\partial\theta\partial x^{a'}} = \frac{\partial f^{dr}}{\partial x^{a'}\partial\theta} = \frac{\partial C^{dr}_{\theta}}{\partial x^{a'}} = 0$$

Such non-integrable relationship between sets of variables is fully treated in Reference 13.

In matrix form the transformation is written

$$dx^k = C^k_{.a}dx^a \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

or

k	
d	dx^{dr}
q	dx^{qr}

$=$

k \ a	a'	b'
d	$\cos x^s$	$-\sin x^s$
q	$\sin x^s$	$\cos x^s$

a	
a'	$dx^{a'}$
b'	$dx^{b'}$

(30*a*)

where x^s is the geometrical variable θ.

The dot in front of the index a indicates that a is a column index. The inverse transformation may be expressed

$$dx^a = C^a_{.k}dx^k$$

or

a	
a'	$dx^{a'}$
b'	$dx^{b'}$

$=$

a \ k	d	q
a'	$\cos x^s$	$\sin x^s$
b'	$-\sin x^s$	$\cos x^s$

k	
d	dx^{dr}
q	dx^{qr}

(31)

Since the relationship is non-integrable and only the differentials of the charges x^k arise in the equations,

$$C^k_{.a} \neq \frac{\partial x^k}{\partial x^a} \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

also, as shown in Appendix 10.3,

$$\frac{\partial C^a_m}{\partial x^n} \neq \frac{\partial C^a_n}{\partial x^m} \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

Such a relationship is non-holonomic.[13] The non-holonomic form of Lagrange's equations was developed by Boltzmann and Hamel.[13] This is written

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{x}^c}\right) - \frac{\partial T}{\partial x^c} + \frac{\partial T}{\partial \dot{x}^d}C^k_c C^n_a\left(\frac{\partial C^d_k}{\partial x^n} - \frac{\partial C^d_n}{\partial x^k}\right)\dot{x}^a + \frac{\partial F}{\partial \dot{x}^c} = f_c \quad . \quad (34)$$

This is seen to consist of Lagrange's eqn. (2) with the addition of terms containing the non-holonomic transformation and its inverse. If the relation between variables is holonomic,

$$\frac{\partial C_k^d}{\partial x^n} - \frac{\partial C_n^d}{\partial x^k} = 0 \quad . \quad . \quad . \quad . \quad . \quad (35)$$

Eqn. (34) may be written in tensor form in a manner similar to eqn. (10), and as shown by Kron[6] and Gibbs[5] it contains terms which retain the non-integrable relations of eqn. (33). The tensor form of the equation is

$$f_c = L_{ca}\frac{d^2x^a}{dt^2} + [ab,c]\frac{dx^a}{dt}\frac{dx^b}{dt} + R_{ca}\frac{dx^a}{dt} \quad . \quad . \quad (36)$$

where now

$$[ab,c] = \frac{1}{2}\left(\frac{\partial L_{cb}}{\partial x^a} + \frac{\partial L_{ca}}{\partial x^b} - \frac{\partial L_{ab}}{\partial x^c}\right) + \Omega_{cb,a} + \Omega_{ca,b} - \Omega_{ab,c} \quad . \quad (37)$$

where
$$\Omega_{ab,c} = \Omega_{ab}^d L_{dc} \quad . \quad . \quad . \quad . \quad . \quad (38)$$

and
$$\Omega_{ab}^d \equiv \tfrac{1}{2}C_a^k C_b^n\left(\frac{\partial C_k^d}{\partial x^n} - \frac{\partial C_n^d}{\partial x^k}\right) \quad . \quad . \quad . \quad . \quad (39)$$

(4) QUASI-HOLONOMIC SYSTEMS[6]

The machine axes transformations in Section 3 are not entirely non-holonomic because the angle θ, the mechanical part of the generalized variable x^k, arises in the analysis and transforms holonomically; i.e. the mechanical part of the transformation is one of variables and not of differentials. This simplifies the analysis considerably.

In transforming from the rotating-axis machine to that with stationary axes the transformation may be expressed

$$\left.\begin{aligned} dx^{ds} &= dx^{ds} \\ dx^{a'} &= \cos x^s\, dx^{dr} + \sin x^s\, dx^{qr} \\ dx^{b'} &= -\sin x^s dx^{dr} + \cos x^s dx^{qr} \\ dx^{qs} &= dx^{qs} \\ dx^s &= dx^s \end{aligned}\right\} \quad . \quad . \quad . \quad (40)$$

or $dx^a = C^a_{.k}dx^k$

$$\begin{array}{c|c} a & \\ \hline ds & dx^{ds} \\ a' & dx^{a'} \\ b' & dx^{b'} \\ qs & dx^{qs} \\ s & dx^s \end{array} = \begin{array}{c|ccccc} a \backslash k & ds & dr & qr & qs & s \\ \hline ds & 1 & & & & \\ a' & & \cos x^s & \sin x^s & & \\ b' & & -\sin x^s & \cos x^s & & \\ qs & & & & 1 & \\ s & & & & & 1 \end{array} \begin{array}{c|c} k & \\ \hline ds & dx^{ds} \\ dr & dx^{dr} \\ qr & dx^{qr} \\ qs & dx^{qs} \\ s & dx^s \end{array} \quad (41)$$

There are two distinct parts of the range of the old variables, namely x^P, the electrical variables, and x^Q, the geometrical variable θ. The L's and R's, whether electrical or mechanical, depend only on x^Q and do not contain the electrical co-ordinates x^P. Such an absence of specific co-ordinates in the components of geometric objects is called in geometry a "cylinder condition."[16]

In a holonomic reference frame the geometric objects are expressed as functions of the co-ordinates x^k, and in a new holonomic system they become functions of new co-ordinates. In a non-holonomic system, however, there exists instead of a set of new co-ordinates, a set of new differentials. Relative to a non-holonomic reference frame the geometric objects are expressed as functions of new differentials, and coefficients, which are functions of the old co-ordinates. In general, therefore, the new equations carry forward the original holonomic co-ordinates in these coefficients, along with the new differentials. In this analysis of machines, however, the geometric objects of the new frame may be expressed without carrying forward the old co-ordinates, because the coefficients are functions only of θ, which transforms holonomically to the same value in the new reference frame.

Thus, even though the new equations have been derived by non-holonomic transformation, there is nothing in their final form to characterize them as being in a non-holonomic reference frame. Therefore, as shown in Section 5, the final equations of the stationary-axis machine may be considered to be holonomic, with a connection $\Gamma_{kn,m}$ which differs in form from that of the first holonomic machine. It is for this reason that a further non-holonomic object may be set up between the machine in reference frame (iii) in Section 5 and that in frame (ii) (considered as holonomic).

The equations of the second primitive machine (with stationary axes) are now written

$$f_m = L_{mk}\frac{d^2x^k}{dt^2} + \left[\frac{1}{2}\left(\frac{\partial L_{mn}}{dx^k} + \frac{\partial L_{mk}}{dx^n} - \frac{\partial L_{kn}}{dx^m}\right) + \Omega_{mn,k} + \Omega_{mk,n} - \Omega_{kn,m}\right]\frac{dx^k}{dt}\frac{dx^n}{dt} + R_{mk}\frac{dx^k}{dt} \quad . \quad (42)$$

where
$$L_{mk} = L_{ac}C_m^a C_k^c \quad . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

$$dx^a = C^a_{.m}dx^m \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

and
$$\Omega_{mn,k} = \tfrac{1}{2}C_m^a C_n^b\left(\frac{\partial C_a^h}{\partial x^b} - \frac{\partial C_b^h}{\partial x^a}\right)L_{hk} \quad . \quad . \quad . \quad (45)$$

From eqns. (41) and (43)

$$L_{mk} = \begin{array}{c|ccccc} m \backslash k & ds & dr & qr & qs & s \\ \hline ds & L_{ds} & M_d & & & \\ dr & M_d & L_{dr} & & & \\ qr & & & L_{qr} & M_q & \\ qs & & & M_q & L_{qs} & \\ s & & & & & L_{ss} \end{array} \quad (46)$$

Thus the terms $\partial L_{kn}/\partial x_m$, etc., in eqn. (42) are zero since the inductances are not functions of either charges or angles.

It is obvious that the term C_a^h may be differentiated only with respect to the mechanical part of the range of variables, x^s, and the expansion of the non-holonomic object is therefore simplified. For example, in eqn. (45), either a or b must be x^s (or θ). Let the values of a, b, m, and n range from 1 to 5 to cover the five rows and columns of C_a^m. The only values of C_a^h and C_b^h inside the bracket of eqn. (45) which can be differentiated to give non-zero values are

$$a = 5, \quad b = 2 \text{ or } 3$$
$$b = 5, \quad a = 2 \text{ or } 3$$

Since either a or b must be 5, the following possible values of $C_m^a C_n^b$ may be written down

$$\begin{array}{cccc} C_5^5C_2^2 & C_5^5C_2^3 & C_2^2C_5^5 & C_2^3C_5^5 \\ C_5^5C_3^2 & C_5^5C_3^3 & C_3^2C_5^5 & C_3^3C_5^5 \end{array}$$

It is further seen by inspection that when a and b are both 5, Ω_{mn}^h is zero, and when a and b are unequal, one being 5, one term of the bracketed quantity is zero.

The general equations are therefore

$$2\Omega^h_{m(s)} = C^a_m C^5_5 \frac{\partial C^h_a}{\partial x^5} \quad . \quad . \quad . \quad . \quad . \quad (47)$$

$$2\Omega^h_{(s)n} = -C^5_5 C^b_n \frac{\partial C^h_b}{\partial x^5} \quad . \quad . \quad . \quad . \quad . \quad (48)$$

The possible non-zero values of Ω^h_{mn} may be written down by inspection and are shown in the following array:

$\Omega^2_{mn} =$

$m \backslash n$	1	2	3	4	5
1					
2					Ω^2_{25}
3					Ω^2_{35}
4					
5		Ω^2_{52}	Ω^2_{53}		

. (49)

A similar array may be written for Ω^3_{mn}. Also

$$2\Omega^2_{(5)3} = C^5_5 C^b_3 \frac{\partial C^2_b}{\partial x^5} \quad (b = 2, 3)$$

$$= \sin^2 x^s + \cos^2 x^s = 1 \quad . \quad . \quad . \quad (50)$$

Similarly

$$\left.\begin{aligned} 2\Omega^2_{3(5)} &= -1 \\ 2\Omega^3_{2(5)} &= 1 \\ 2\Omega^3_{5(2)} &= -1 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (51)$$

All the other values of Ω^h_{mn} are zero.

The resulting object is $\Omega^h_{m(s)}$ or $\Omega^h_{(s)n}$ according to whether m or n takes the geometrical angle co-ordinate value, x^s:

$2\Omega^h_{(s)n} =$

$h \backslash n$	ds	dr	qr	qs	s
ds					
dr			1		
qr		-1			
qs					
s					

. (52*a*)

$2\Omega^h_{m(s)} =$

$h \backslash m$	ds	dr	qr	qs	s
ds					
dr			-1		
qr		1			
qs					
s					

. (52*b*)

The object Ω^h_{mn} is seen to be skew-symmetric in the indices m and n.

In the equation of Lagrange the "Christoffel symbol" is written $[ab,c]$. In the non-holonomic form of the equation the corresponding connection is

$$[uv,w] + \Omega_{wv,u} + \Omega_{wu,v} - \Omega_{uv,w} \quad . \quad . \quad . \quad (53)$$

This is made up of the symmetrical part $[uv,w]$ and the skew-symmetric part $(\Omega_{wv,u} + \Omega_{wu,v} - \Omega_{uv,w})$, these together making an asymmetrical connection[2] written $\Gamma_{uv,w}$.

The non-holonomic equation of motion of the machine is therefore written as in eqn. (25):

$$e_w = R_{wv}\frac{dx^v}{dt} + L_{wv}\frac{d^2x^v}{dt^2} + \Gamma_{uv,w}\frac{dx^u}{dt}\frac{dx^v}{dt} \quad . \quad . \quad (54)$$

The electrical and mechanical parts of the equation are shown in Section 2, eqns. (26) and (27).

The symmetrical part is zero since L_{wv} is independent of x^s (which is equivalent to θ). The connection has then the following components:

$$\Gamma_{uv,w}i^u i^v \equiv \Gamma_{sn,m}i^s i^n + \Gamma_{ks,m}i^k i^s + \Gamma_{kn,s}i^k i^n. \quad . \quad (55)$$

where, for example,

$$\Gamma_{sn,m}i^s i^n \equiv (\Omega_{mn,s} + \Omega_{ms,n} - \Omega_{sn,m})i^s i^n \quad . \quad . \quad (56)$$

where $$i^s \equiv \frac{d\theta}{dt}$$

The term $\Omega_{mn,s}$ is seen from eqns. (45) and (48) to be zero, and since $\Omega_{sn,m}$ is skew symmetric in the indices s and n, the term $\Omega_{sn,m}i^s i^n$ is also zero.

Therefore

$$\Gamma_{sn,m}i^s i^n = \Omega_{ms,n}i^s i^n \quad . \quad . \quad . \quad . \quad . \quad (57)$$

$$\Gamma_{ks,m}i^k i^s = \Omega_{ms,k}i^k i^s \quad . \quad . \quad . \quad . \quad . \quad (58)$$

$$\Gamma_{kn,s}i^k i^n = 2\Omega_{sn,k}i^k i^n \quad . \quad . \quad . \quad . \quad . \quad (59)$$

Thus the electrical equation becomes

$$e_m = R_{mn}i^n + L_{mn}\frac{di^n}{dt} + 2\Omega_{ms,k}i^s i^k \quad . \quad . \quad . \quad (60)$$

and the mechanical equation is

$$f_s = R_{st}i^t + L_{st}\frac{di^t}{dt} + 2\Omega_{sn,k}i^n i^k \quad . \quad . \quad . \quad (61)$$

The last terms in eqns. (60) and (61) are found on expansion, using eqns. (46) and (52), to be

$2\Omega_{ms,k} =$

$m \backslash k$	ds	dr	qr	qs	s
ds					
dr			L_{qr}	M_q	
qr	$-M_d$	$-L_{dr}$			
qs					
s					

$= -2\Omega_{sn,k}$ (62)

Eqn. (60) is now seen to be identical with the matrix eqn. (20) written down by inspection in Section 2. The equations for the commutator or stationary-axis machine have been derived by transformation of the dynamical equation of motion of the holonomic rotating-axis or slip-ring machine.

If now a further transformation be carried out to axes rotating at any arbitrary angular velocity $p\theta_1$ it is found that some of the terms of the equation transform tensorially while additional voltage terms arise which are due only to the measurements from the new reference frame. Tensor analysis separates the tensor quantities which arise in all reference frames, from those quantities which may arise in one reference frame and disappear in others. The nature of the connection $\Gamma_{uv,w}$ is examined by Eqn. (65) is shown in Section 2, eqn. (7). The inductance tensor is shown in matrix 9. From the inductance matrix and eqn. (11),

$$[ab,c]i^a i^b \equiv [sb,c]i^s i^b + [as,c]i^a i^s = 2[as,c]i^a i^s$$

and

$$2[as,c]i^a i^s = \frac{\partial L_{ca}}{\partial \theta}\frac{d\theta}{dt}i^a$$

From matrix 9, $\dfrac{\partial L_{ca}}{\partial \theta}$ is

$c \backslash a$	ds	a'	b'	qs
ds		$-M_d \sin\theta$	$-M_d \cos\theta$	
a'	$-M_d \sin\theta$	$2(L_{qr} - L_{dr}) \sin\theta \cos\theta$	$(L_{qr} - L_{dr})(\cos^2\theta - \sin^2\theta)$	$M_q \cos\theta$
b'	$-M_d \cos\theta$	$(L_{qr} - L_{dr})(\cos^2\theta - \sin^2\theta)$	$-2(L_{qr} - L_{dr}) \cos\theta \sin\theta$	$-M_q \sin\theta$
qs		$M_q \cos\theta$	$-M_q \sin\theta$	

. . . (67)

The equation is therefore

$$e_c = R_{ca}i^a + L_{ca}pi^a + \frac{\partial L_{ca}}{\partial \theta}\frac{d\theta}{dt}i^a \quad . . . . . (68a)$$

This may be written

$$e_c = R_{ca}i^a + L_{ca}pi^a + V_{ca}i^a p\theta + G_{ca}i^a p\theta \quad . . . . . (68b)$$

where

$V_{ca} =$

$c \backslash a$	ds	a'	b'	qs
ds		$-M_d \sin\theta$	$-M_d \cos\theta$	
a'		$(L_{qr} - L_{dr}) \sin\theta \cos\theta$	$-L_{qr} \sin^2\theta - L_{dr} \cos^2\theta$	
b'		$L_{qr} \cos^2\theta + L_{dr} \sin^2\theta$	$-(L_{qr} - L_{dr}) \sin\theta \cos\theta$	
qs		$M_q \cos\theta$	$-M_q \sin\theta$	

. . . (68c)

$G_{ca} =$

$c \backslash a$	ds	a'	b'	qs
ds				
a'	$-M_d \sin\theta$	$(L_{qr} - L_{dr}) \sin\theta \cos\theta$	$L_{dr} \sin^2\theta + L_{qr} \cos^2\theta$	$M_q \cos\theta$
b'	$-M_d \cos\theta$	$-L_{dr} \cos^2\theta - L_{qr} \sin^2\theta$	$-(L_{qr} - L_{dr}) \sin\theta \cos\theta$	$-M_q \sin\theta$
qs				

. . . (68d)

considering the transformation to general rotating axes. The objects $\Gamma_{uv,w}$ and Ω^w_{uv} have the following laws of transformation:[2,5]

$$\Gamma_{u'v',w'} = \Gamma_{uv,w}C^u_{u'}C^v_{v'}C^w_{w'} + L_{wu}C^w_{w'}\frac{\partial C^u_{u'}}{\partial x^{v'}} \quad . . (63)$$

and

$$\Omega^{w'}_{u'v'} = \Omega^w_{uv}C^u_{u'}C^v_{v'}C^{w'}_w + \frac{1}{2}\left(\frac{\partial C^{w'}_u}{\partial x^v} - \frac{\partial C^{w'}_v}{\partial x^u}\right)C^u_{u'}C^v_{v'} \quad . (64)$$

The quantities $\Gamma_{uv,w}$ and $\Omega_{uv,w}$ are therefore not tensors.

The equations of the electrical machine with general rotating axes are derived in Section 5.

(5) THREE REFERENCE FRAMES

(5.1) Electrical Equations

(5.1.1) The Holonomic Frame. (i)

$$e_c = R_{ca}i^a + pL_{ca}i^a \quad . . . . . (65)$$

$$e_c = R_{ca}i^a + L_{ca}pi^a + [ab,c]i^a i^b \quad . . . . (66)$$

(5.1.2) The Stationary Axis Frame. (ii)

From Fig. 2 it is seen that the transformation from frame (i) to frame (ii) is given by

$$dx^c = C^c_{\cdot m}dx^m$$

where

$C^c_{\cdot m} =$

$c \backslash m$	ds	dr	qr	qs
ds	1			
a'		$\cos\theta$	$\sin\theta$	
b'		$-\sin\theta$	$\cos\theta$	
qs				1

. . . (69)

The equations for frame (ii) are

$$e_m = R_{mk}i^k + L_{mk}pi^k + \Gamma_{kn,m}i^k i^n \quad . . . (70)$$

where

$$\Gamma_{kn,m}i^k i^n \equiv \Gamma_{sn,m}i^s i^n + \Gamma_{ks,m}i^k i^s \quad . . . (71)$$

Also $\Gamma_{kn,m}i^k i^n = [kn,m]i^k i^n + (\Omega_{mn,k} + \Omega_{mk,n} - \Omega_{kn,m})i^k i^n$
$= 2\Omega_{ms,k}i^s i^k$. (72)

and $L_{\gamma\alpha} = L_{ca}C^c_\gamma C^a_\alpha$ . . . . . (79)

This becomes

	ds	S_1	S_2	qs
ds	L_{ds}	$M_d \cos\theta_1$	$-M_d \sin\theta_1$	
S_1	$M_d \cos\theta_1$	$L_{dr}\cos^2\theta_1 + L_{qr}\sin^2\theta_1$	$(L_{qr} - L_{dr})\sin\theta_1\cos\theta_1$	$M_q \sin\theta_1$
S_2	$-M_q \sin\theta_1$	$(L_{qr} - L_{dr})\sin\theta_1\cos\theta_1$	$L_{dr}\sin^2\theta_1 + L_{qr}\cos^2\theta_1$	$M_q \cos\theta_1$
qs		$M_q \sin\theta_1$	$M_q \cos\theta_1$	L_{qs}

(80)

The terms L_{mk} and $2\Omega_{ms,k}$ are shown in Section 4, matrices 46 and 62. The equations may therefore be written

$$e_m = R_{mk}i^k + L_{mk}pi^k + G_{mk}i^k p\theta \quad . \quad . \quad . \quad (73)$$

(5.1.3) The Frame having Rotor Axes Rotating Freely. (iii)

The transformation matrix from frame (i) to frame (iii) is

$C^a_\alpha =$

$a \backslash \alpha$	ds	S_1	S_2	qs
ds	1			
a'		$\cos\theta_f$	$\sin\theta_f$	
b'		$-\sin\theta_f$	$\cos\theta_f$	
qs				1

(74)

where, from Fig. 3, $\theta_f = \theta - \theta_1$.

The geometrical variable x^s transforming holonomically remains θ, from matrix eqn. (40).

Now $$[\alpha\beta,\gamma] = \frac{1}{2}\left(\frac{\partial L_{\gamma\beta}}{\partial x^\alpha} + \frac{\partial L_{\gamma\alpha}}{\partial x^\beta} - \frac{\partial L_{\alpha\beta}}{\partial x^\gamma}\right) \quad . \quad . \quad (81)$$

and $$[s\beta,\gamma]i^s i^\beta + [\alpha s,\gamma]i^\alpha i^s = 2[\alpha s,\gamma]i^\alpha i^s \quad . \quad . \quad . \quad (82)$$

and
$$2[\alpha s,\gamma]i^\alpha i^s = \left(\frac{\partial L_{\gamma s}}{\partial x^\alpha} + \frac{\partial L_{\gamma\alpha}}{\partial x^s} - \frac{\partial L_{\alpha s}}{\partial x^\gamma}\right)i^\alpha p\theta$$
$$= \frac{\partial L_{\gamma\alpha}}{\partial\theta}i^\alpha p\theta$$
$$= \frac{\partial L_{\gamma\alpha}}{\partial\theta_1}\frac{d\theta_1}{d\theta}\frac{d\theta}{dt}i^\alpha$$

Therefore $$2[\alpha s,\gamma]i^\alpha i^s = \frac{\partial L_{\gamma\alpha}}{\partial\theta_1}\frac{d\theta_1}{dt}i^\alpha \quad . \quad . \quad . \quad . \quad (83)$$

$\frac{\partial L_{\gamma\alpha}}{\partial\theta_1} =$

$\gamma \backslash \alpha$	ds	S_1	S_2	qs
ds		$-M_d \sin\theta_1$	$-M_d \cos\theta_1$	
S_1	$-M_d \sin\theta_1$	$2(L_{qr} - L_{dr})\sin\theta_1\cos\theta_1$	$L_{qr}\cos^2\theta_1 - L_{qr}\sin^2\theta_1$ $-L_{dr}\cos^2\theta_1 + L_{dr}\sin^2\theta_1$	$M_q \cos\theta_1$
S_2	$-M_d \cos\theta_1$	$L_{qr}\cos^2\theta_1 - L_{qr}\sin^2\theta_1$ $-L_{dr}\cos^2\theta_1 + L_{dr}\sin^2\theta_1$	$-2(L_{qr} - L_{dr})\cos\theta_1\sin\theta_1$	$-M_q \sin\theta_1$
qs		$M_q \cos\theta_1$	$-M_q \sin\theta_1$	

(84)

The equations in frame (iii) are

$$e_\gamma = R_{\gamma\alpha}i^\alpha + L_{\gamma\alpha}pi^\alpha + \Gamma_{\alpha\beta,\gamma}i^\alpha i^\beta \quad . \quad . \quad . \quad (75)$$

where $$\Gamma_{\alpha\beta,\gamma}i^\alpha i^\beta \equiv \Gamma_{s\beta,\gamma}i^s i^\beta + \Gamma_{\alpha s,\gamma}i^\alpha i^s + \Gamma_{\alpha\beta,s}i^\alpha i^\beta \quad . \quad (76)$$

and $$\Gamma_{\alpha\beta,\gamma}i^\alpha i^\beta = \{[\alpha\beta,\gamma] + \Omega_{\gamma\beta,\alpha} + \Omega_{\gamma\alpha,\beta} - \Omega_{\alpha\beta,\gamma}\}i^\alpha i^\beta \quad . \quad (77)$$

The electrical part of the equation therefore becomes

$$e_\gamma = R_{\gamma\alpha}i^\alpha + L_{\gamma\alpha}pi^\alpha + \{2[\alpha s,\gamma]i^\alpha i^s + 2\Omega_{\gamma s,\alpha}i^\alpha i^s\} \quad . \quad (78)$$

From eqns. (38) and (47)

$$2\Omega_{\gamma s,\alpha}i^s i^\alpha = \frac{\partial C^{\cdot\delta}_c}{\partial\theta}C^{\cdot c}_\gamma L_{\delta\alpha}\frac{d\theta}{dt}i^\alpha$$
$$= \frac{\partial C_t^{-1}}{\partial\theta}C_t L_{\delta\alpha}\frac{d\theta}{dt}i^\alpha \quad . \quad . \quad . \quad . \quad (85)$$

Therefore $$2\Omega_{\gamma s,\alpha}i^s i^\alpha = \frac{\partial C_t^{-1}}{\partial\theta_f}C_t L_{\delta\alpha}\frac{d\theta_f}{dt}i^\alpha \quad . \quad . \quad . \quad . \quad (86)$$

where C_t is the transpose of matrix C^c_γ (matrix 74). $2\Omega_{\gamma s,\alpha}$ is then

$\gamma \backslash \alpha$	ds	S_1	S_2	qs
ds				
S_1	$-M_d \sin\theta_1$	$(L_{qr} - L_{dr})\sin\theta_1\cos\theta_1$	$L_{dr}\sin^2\theta_1 + L_{qr}\cos^2\theta_1$	$M_q \cos\theta_1$
S_2	$-M_d \cos\theta_1$	$-L_{dr}\cos^2\theta_1 - L_{qr}\sin^2\theta_1$	$-(L_{qr} - L_{dr})\sin\theta_1\cos\theta_1$	$-M_q \sin\theta_1$
qs				

(87)

The equations may therefore be written

$$e_\gamma = R_{\gamma\alpha}i^\alpha + L_{\gamma\alpha}pi^\alpha + V_{\gamma\alpha}i^\alpha p\theta_1 + G_{\gamma\alpha}i^\alpha p\theta + G_{\gamma\alpha}i^\alpha p\theta_f \quad . \quad (88)$$

or

$$e_\gamma = R_{\gamma\alpha}i^\alpha + L_{\gamma\alpha}pi^\alpha + V_{\gamma\alpha}i^\alpha p\theta_1 + G_{\gamma\alpha}i^\alpha p\theta \quad . \quad . \quad (89)$$

The equations of the machine in reference frame (iii) may also be obtained by transformation of the reference frame (ii).

From Fig. 3 the transformation matrix is

$C^m_\gamma =$

$m \backslash \gamma$	ds	S_1	S_2	qs
ds	1			
dr		$\cos\theta_1$	$-\sin\theta_1$	
qr		$\sin\theta_1$	$\cos\theta_1$	
qs				1

. . (90)

$$L_{\gamma\alpha} = L_{mk}C^m_\gamma C^k_\alpha = \text{matrix 80} \quad . \quad . \quad . \quad (91)$$

The equations may be written

$$e_\gamma = R_{\gamma\alpha}i^\alpha + L_{\gamma\alpha}pi^\alpha + \Gamma_{\alpha\beta,\gamma}i^\alpha i^\beta \quad . \quad . \quad . \quad (92)$$

where

$$\Gamma_{\alpha\beta,\gamma} \equiv [\alpha\beta,\gamma] + \Omega_{\gamma\beta,\alpha} + \Omega_{\gamma\alpha,\beta} - \Omega_{\alpha\beta,\gamma} \quad . \quad (92a)$$

The terms $\Omega_{\alpha\beta,\gamma}$ are non-holonomic objects of the frame (iii) with respect to the holonomic frame (i), and

$$\Omega_{\alpha\beta,\gamma} = \Omega_{kn,m}C^k_\alpha C^n_\beta C^m_\gamma + \frac{1}{2}\left(\frac{\partial C^\delta_k}{\partial x^n} - \frac{\partial C^\delta_n}{\partial x^k}\right)C^k_\alpha C^n_\beta L_{\delta\gamma} \quad . \quad (93)$$

Now

$$\Gamma_{\alpha\beta,\gamma}i^\alpha i^\beta = \{2[\alpha s,\gamma] + 2\Omega_{\gamma s,\alpha}\}i^\alpha p\theta \quad . \quad . \quad (94)$$

Also

$$2\Omega_{\gamma s,\alpha} = 2\Omega_{ms,k}C^m_\gamma C^s_s C_k + \frac{\partial C^\delta_k}{\partial x^s}C^k_\alpha C^s_s L_{\delta\gamma} \quad . \quad . \quad (95)$$

and $2\Omega'_{\gamma s,\alpha} =$ matrix 87

where

$$\Omega'_{\gamma s,\alpha} \equiv \Omega_{ms,k}C^m_\gamma C^s_s C^k_\alpha \quad . \quad . \quad . \quad . \quad (96)$$

And

$$\frac{\partial C^\delta_k}{\partial\theta}C^k L_{\delta\gamma}p\theta = \frac{\partial C^\delta_k}{\partial\theta_1}C^k_\alpha L_{\delta\gamma}p\theta_1$$

$$\frac{\partial C^\delta_k}{\partial\theta_1}C^k_\alpha L_{\delta\gamma} = \text{minus matrix 87} \quad . \quad . \quad . \quad (97)$$

$$2[\alpha s,\gamma] = \text{matrix 84} \quad . \quad . \quad . \quad . \quad . \quad (98)$$

The equations therefore may be written

$$e_\gamma = R_{\gamma\alpha}i^\alpha + L_{\gamma\alpha}pi^\alpha + V_{\gamma\alpha}i^\alpha p\theta_1 + G_{\gamma\alpha}i^\alpha p\theta_1 + G_{\gamma\alpha}i^\alpha p\theta - G_{\gamma\alpha}i^\alpha p\theta_1 \quad . \quad (99)$$

or

$$e_\gamma = R_{\gamma\alpha}i^\alpha + L_{\gamma\alpha}pi^\alpha + V_{\gamma\alpha}i^\alpha p\theta_1 + G_{\gamma\alpha}i^\alpha p\theta \quad . \quad (100)$$

(5.2) Equations of Torque

(5.2.1) Frame (i).

$$f_s = R_{st}\frac{d\theta}{dt} + L_{st}\frac{d^2\theta}{dt^2} + [ab,s]i^a i^b \quad . \quad . \quad (101)$$

or

$$f_s = R_{st}\frac{d\theta}{dt} + L_{st}\frac{d^2\theta}{dt^2} - \frac{1}{2}\frac{\partial L_{ab}}{\partial\theta}i^a i^b \quad . \quad . \quad (102)$$

An asymmetrical matrix A may be written as the sum of symmetrical and skew-symmetrical parts, thus[2]

$$A = \frac{A + A_t}{2} + \frac{A - A_t}{2}$$

Using this relationship it may be shown that

$$\frac{1}{2}\frac{\partial L_{ab}}{\partial\theta}i^a i^b = G_{ab}i^a i^b \quad . \quad . \quad . \quad . \quad . \quad (103)$$

thus

$$f_s = R_{st}\frac{d\theta}{dt} + L_{st}\frac{d^2\theta}{dt^2} - G_{ab}i^a i^b \quad . \quad . \quad . \quad (104)$$

(5.2.2) Frame (ii).

$$f_u = R_{uv}\frac{d\theta}{dt} + L_{uv}\frac{d^2\theta}{dt^2} - G_{nk}i^n i^k \quad . \quad . \quad . \quad (105)$$

(5.2.3) Frame (iii).

$$f_s = R_{st}\frac{d\theta}{dt} + L_{st}\frac{d^2\theta}{dt^2} + [\alpha\beta,s]i^\alpha i^\beta + 2\Omega_{s\beta,\alpha}i^\alpha i^\beta \quad . \quad (106)$$

$$f_s = R_{st}\frac{d\theta}{dt} + L_{st}\frac{d^2\theta}{dt^2} - \frac{1}{2}\frac{\partial L_{\alpha\beta}}{\partial\theta}i^\alpha i^\beta - G'_{\alpha\beta}i^\alpha i^\beta \quad . \quad (106a)$$

Now

$$\frac{1}{2}\frac{\partial L_{\alpha\beta}}{\partial\theta}i^\alpha i^\beta = \frac{1}{2}\frac{\partial L_{\alpha\beta}}{\partial\theta_1}\frac{d\theta_1}{d\theta}i^\alpha i^\beta \quad . \quad . \quad . \quad (107)$$

and

$$G'_{\alpha\beta} = \frac{\partial C_t^{-1}}{\partial\theta_f}C_t\frac{d\theta_f}{d\theta}L_{\delta\alpha} \quad . \quad . \quad . \quad . \quad (107a)$$

Thus

$$f_s = R_{st}\frac{d\theta}{dt} + L_{st}\frac{d^2\theta}{dt^2} - G_{\alpha\beta}\left(\frac{d\theta_1}{d\theta} + \frac{d\theta_f}{d\theta}\right)i^\alpha i^\beta \quad . \quad (108)$$

or

$$f_s = R_{st}\frac{d\theta}{dt} + L_{st}\frac{d^2\theta}{dt^2} - G_{\alpha\beta}i^\alpha i^\beta \quad . \quad . \quad . \quad (109)$$

(5.3) The Torque Tensor

It is seen from eqns. (62) and (87) that

$$G_{\gamma\alpha} = G_{mk}C^m_\gamma C^k_\alpha \quad . \quad . \quad . \quad . \quad . \quad (110)$$

It is also found on examination of the matrices in the previous Sections that

$$G_{mk} = G_{ca}C^c_m C^a_k \quad . \quad . \quad . \quad . \quad . \quad (111)$$

and

$$G_{\gamma\alpha} = G_{ca}C^c_\gamma C^a_\alpha \quad . \quad . \quad . \quad . \quad . \quad (112)$$

The torque matrix therefore transforms as a tensor and is associated with the holonomic variable θ_i. The equations of frame (ii) may therefore be written[6]

$$e_m = R_{mk}i^k + L_{mk}pi^k + T_{msk}i^s i^k \quad . \quad . \quad (113)$$

where T_{msk} is a tensor; or

$$e_m = R_{mk}i^k + L_{mk}pi^k + (-S_{mnk} - S_{mkn} + S_{knm})i^k i^n \quad . \quad (114)$$

where S_{knm} is defined as a tensor having components equal to, but the negative of, $\Omega_{kn,m}$. The negative value is chosen in order that the tensor here defined will be that given by

$$\tfrac{1}{2}(\Gamma_{kn,m} - \Gamma_{nk,m}) \quad . \quad . \quad . \quad . \quad . \quad (115)$$

the skew-symmetric part of $\Gamma_{kn,m}$. That this is a tensor may be proved by the equations of tensor calculus.[14] This is a well-known tensor quantity in geometry of n-dimensional spaces and is there termed the "torsion" tensor. In terms of the torsion tensor the equations of frame (iii), derived from frame (ii), may be written

$$e_\gamma = R_{\gamma\alpha}i^\alpha + L_{\gamma\alpha}pi^\alpha + \{[\alpha\beta,\gamma] - S_{\gamma\beta\alpha} - S_{\gamma\alpha\beta} + S_{\alpha\beta\gamma} + \Omega_{\gamma\beta,\alpha} + \Omega_{\gamma\alpha,\beta} - \Omega_{\alpha\beta,\gamma}\}i^\alpha i^\beta \quad . \quad (116)$$

where

$$\Omega_{\alpha\beta,\gamma} = \left(\frac{\partial C^\delta_k}{\partial x^n} - \frac{\partial C^\delta_n}{\partial x^k}\right)C^k_\alpha C^n_\beta L_{\delta\gamma} \quad . \quad . \quad . \quad (117)$$

The terms $(-S_{\gamma\beta\alpha}S_{\gamma}{}_{\alpha\beta} + S_{\alpha\beta\gamma})$ and $(\Omega_{\gamma\beta,\alpha} + \Omega_{\gamma\alpha,\beta} - \Omega_{\alpha\beta,\gamma})$ may be compared with eqn. (93) where on the right-hand side there are a tensor term and a non-holonomic object.

It is thus shown that in any of these three frames the following terms transform as tensors, e, i, L and G; also, G is associated with $p\theta$ the rotor speed. The generated voltage term due to V arises because of the choice of reference frame and is a function of the angular velocity of the frame with respect to the direct axis. If the angular velocity $p\theta_1$ becomes zero, θ_f becomes θ and the equations of frame (iii) give those of frame (ii). If $p\theta_1$ becomes $p\theta$, θ_f becomes zero and the frame (iii) equations give those of frame (i).

These relationships may also be derived by transforming Γ as a whole[2] using eqn. (63), instead of transforming the components of Γ as has been done here. This method, however, does not show so clearly the mechanism of the transformations.

(6) APPLICATION

The tensor equations of electrical machines may be used in two ways.[1] In the first, a comparison is made between the primitive machine and another type of machine whose equations are required; for example, a metadyne, Schrage motor, etc. This aspect of Kron's work has been extensively treated.[2,5,10,11] In such analysis the connection matrix C is set up between the currents in the primitive machine coils and those in the interconnected coils of the derived machine, and the equations of the primitive machine are then transformed as shown by Kron, Gibbs and others, to give the required equations. The second application has been used more recently.[20] This consists of transformations among the reference frames of a given machine.[15] A familiar example is that of the synchronous alternator which may be analysed by setting up the equations with respect to quantities appearing at stationary 3-phase armature terminals, or alternatively by using Park's equations which contain quantities arising with respect to rotating direct- and quadrature-field axes. The d- and q-axes quantities are, of course, fictitious, but this reference frame leads to linear differential equations, often with constant coefficients, and is therefore widely used. It has been found,[20,21] however, that Park's form of the equations becomes very complicated when rotor oscillations occur in the machine. In hunting studies a more suitable frame is one which rotates freely at synchronous angular velocity and is independent of the rotor oscillations. The equations in this form, of course, become identical with Park's equations when the rotor has a uniform angular velocity with no oscillations or acceleration. It is not proposed to discuss the oscillations of machines here, nor indeed to deal in detail with the use of the synchronous-machine equations, but simply to present the concepts of changes of reference frames using the tensor technique developed by Kron.

The reason for changes of reference frame is apparent from examination of Park's equations. In a rotating-field alternator the d- and q-axes of reference rotate synchronously with the field structure relative to the armature. Under balanced steady-state conditions the fluxes and current and voltage vectors along these axes will be constant in magnitude and rotating in space. Thus the steady-state equations of the alternator are obtained from eqns. (21) by putting terms such as $L_d p$ equal to zero, where p operates on a current component, retaining the p terms where p operates only on θ to give the angular velocity. The steady-state equations are then

$$\left.\begin{aligned} e_f &= R_f i^f \\ e_d &= -R_d i^{dr} + L_q p\theta i^{qr} \\ e_q &= -R_q i^{qr} - M_d p\theta i^f - L_d p\theta i^{dr} \end{aligned}\right\} \quad . \quad . \quad (118)$$

It is obvious that if differential equations are to be set up for any machine or network to which the alternator is connected, these must be expressed along the same reference frame, the operator p must have the same significance and the steady-state equations must therefore be obtained as before when terms such as Lp become zero. In most equations of a.c. machines and networks, with sinusoidal voltage and currents, the steady-state equations are obtained when Lp becomes $j\omega L$. In this case a transformation of reference frame is required if these machines and networks are to be analysed in conjunction with interconnected synchronous machines. Two very simple cases will illustrate the required transformations, namely the equations of a simple series impedance having resistance and inductive reactance, and those of a 3-phase induction motor. Both of these have been analysed by Kron, but the analysis as set out below demonstrates details of the general method of using tensor equations for this purpose. The transient equation of a simple RL series impedance may be written

$$e = Ri + Lpi$$

Under steady-state conditions, with sinusoidal voltage applied, the equation becomes

$$e = Ri + j\omega Li$$

which may be obtained from the transient equation by putting p equal to $j\omega$. When a 3-phase system is being considered the instantaneous line currents and phase voltages and impedances may be resolved into Clarke components.[22] In order to conform to the phase positions and direction of rotation shown in Fig. 1, the current components may be defined as follows:

$$i^{b'} = \tfrac{1}{3}(2i^A - i^B - i^C) \quad . \quad . \quad . \quad . \quad (119)$$

$$i^{a'} = \frac{1}{\sqrt{3}}(i^B - i^C) \quad . \quad . \quad . \quad . \quad . \quad (120)$$

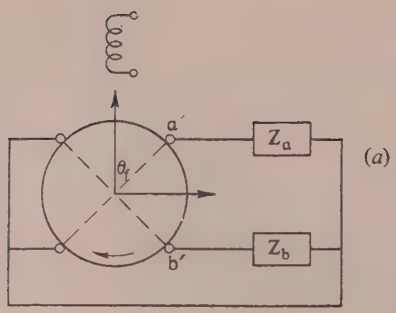
$$i^0 = \tfrac{1}{3}(i^A + i^B + i^C) \quad . \quad . \quad . \quad . \quad (121)$$

where i^A, i^B, i^C are the instantaneous line currents (Miss Clarke uses indices α and β instead of b' and a' as written here). Zero-sequence currents i^0 are those residual currents in the neutral connection to an unbalanced load or point of fault. To simplify the analysis a balanced system will be considered with no zero-sequence currents. In a machine wound for three phases these instantaneous components $i^{b'}$ and $i^{a'}$ lie respectively along the axis of phase A and along the common axis of phases B and C in quadrature with phase A. These are the same as the axes b' and a' used in the holonomic primitive machine, Fig. 1, and are stationary with respect to the armature phase windings. The a' and b' components of an external 3-phase network would be connected to the machine axes as shown in Fig. 5(*a*). It is therefore possible to define for either a machine or stationary network a set of currents i^{S1} and i^{S2} expressed along axes rotating with uniform angular velocity with respect to the axes of the Clarke components. From Fig. 3 the relationships among such currents may be written

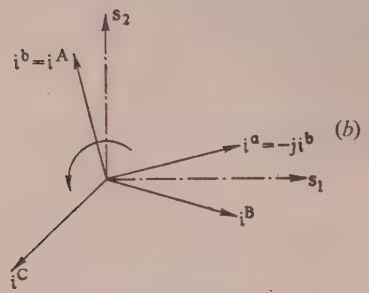
$$\left.\begin{aligned} i^{a'} &= i^{S1}\cos\theta_f + i^{S2}\sin\theta_f \\ i^{b'} &= -i^{S1}\sin\theta_f + i^{S2}\cos\theta_f \end{aligned}\right\} \quad . \quad . \quad . \quad (121)$$

For a stationary network the holonomic (Lagrangian) equations in terms of a' and b' components are

$$\begin{array}{c|c} w & \\ \hline a' & e_{a'} \\ \hline b' & e_{b'} \end{array} = \begin{array}{c|cc} w\backslash v & a' & b' \\ \hline a' & R + Lp & \\ \hline b' & & R + Lp \end{array} \begin{array}{c|c} & \\ \hline a' & i^{a'} \\ \hline b' & i^{b'} \end{array} \quad (122)$$



(*a*)



(*b*)

Fig. 5.—(*a*) Synchronous machine with external network. (*b*) Clarke components of a balanced 3-phase system.

or in index notation

$$e_w = R_{wv}i^v + L_{wv}pi^v$$

where

$R_{wv} =$

w \ v	a'	b'
a'	R	
b'		R

and $L_{wv} =$

w \ v	a'	b'
a'	L	
b'		L

In the free frame the equations become [see eqn. (78)]

$$e_\gamma = R_{\gamma\alpha}i^\alpha + L_{\gamma\alpha}pi^\alpha + 2[\alpha s,\gamma]i^s i^\alpha + 2\Omega_{\gamma s,\alpha}i^s i^\alpha \quad . \quad (123)$$

where [from eqn. (121)]

$C^v_{\cdot\alpha} =$

v \ α	S_1	S_2
a'	$\cos\theta_f$	$\sin\theta_f$
b'	$-\sin\theta_f$	$\cos\theta_f$

. . . (124)

and $$i^v = C^v_{\cdot\alpha}i^\alpha$$

$$e_\alpha = C^v_{\cdot\alpha}e_v \quad (C^v_{\cdot\alpha} \text{ is the transpose of } C^v_{\cdot\alpha})$$

$$R_{\gamma\alpha} = R_{wv}C^w_\gamma C^v_\alpha$$

$$L_{\gamma\alpha} = L_{wv}C^w_\gamma C^v_\alpha$$

It is found that in this case

$$R_{\gamma\alpha} = R_{wv}$$

$$L_{\gamma\alpha} = L_{wv}$$

Thus $$2[\gamma s,\alpha] \equiv \left(\frac{\partial L_{\alpha s}}{\partial x^\gamma} + \frac{\partial L_{\alpha\gamma}}{\partial x^s} - \frac{\partial L_{\gamma s}}{\partial x^\alpha}\right) = 0$$

and therefore $$e_\gamma = R_{\gamma\alpha}i^\alpha + L_{\gamma\alpha}pi^\alpha + 2\Omega_{\gamma s,\alpha}i^s i^\alpha \quad . \quad . \quad (125)$$

where $$2\Omega_{\gamma s,\alpha}i^s i^\alpha = 2\Omega^\delta_{\gamma s}L_{\delta\alpha}i^\alpha p\theta = \frac{\partial C^\delta_w}{\partial\theta_f}C^w_\gamma L_{\delta\alpha}i^\alpha p\theta_f \quad . \quad (126)$$

In direct notation $$C^v_{\cdot\alpha} \equiv \boldsymbol{C} \text{ and } 2\Omega^\delta_{\gamma s} = \boldsymbol{C}_t\frac{\partial \boldsymbol{C}_t^{-1}}{\partial\theta_f} \quad . \quad . \quad (127)$$

Thus $2\Omega^\delta_{\gamma s} =$

γ \ δ	S_1	S_2
S_1		$+1$
S_2	-1	

. . . . (128)

and $2\Omega_{\gamma s,\alpha} =$

γ \ α	S_1	S_2
S_1		$+L$
S_2	$-L$	

$= V_{\gamma\alpha}$ . . (129)

Thus $$e_\gamma = R_{\gamma\alpha}i^\alpha + L_{\gamma\alpha}pi^\alpha + V_{\gamma\alpha}i^\alpha p\theta_f \quad . \quad . \quad . \quad (130)$$

A rotation matrix may be defined by[24]

$$V_{\gamma\alpha} = \rho^\delta_\gamma L_{\delta\alpha} \quad . \quad . \quad . \quad . \quad . \quad (131)$$

where $\rho^\delta_\gamma = 2\Omega^\delta_{\gamma s} =$

γ \ δ	S_1	S_2
S_1		$+1$
S_2	-1	

. . . (132)

This is the rotation matrix used by Kron.[24]

In eqns. (132) above it is not a tensor, being part of the non-holonomic object $\Omega_{\gamma\beta\alpha}$. A similar matrix arises in connection with the formation of the torsion tensor defined in Section 5, and this is then referred to as the "rotation tensor" (there is an algebraic connection between this tensor and the coefficient of rotation defined by Ricci[14]).

The equations of the 3-phase network in terms of axes analogous to the rotating d and q axes of Park thus become

γ	
S_1	e_{S1}
S_2	e_{S2}

$=$

γ \ α	S_1	S_2
S_1	$R + Lp$	$Lp\theta_f$
S_2	$-Lp\theta_f$	$R + Lp$

α	
S_1	i^{S1}
S_2	i^{S2}

(133)

where
$$\left.\begin{aligned} e_{S1} &= e_{a'}\cos\theta_f - e_{b'}\sin\theta_f \\ e_{S2} &= e_{a'}\sin\theta_f + e_{b'}\cos\theta_f \end{aligned}\right\} \quad . \quad . \quad . \quad (134)$$

By trigonometrical substitution it may be shown that in the steady state, when $e_{b'} = E_A\cos\omega t$ and $e_{b'} = je_{a'}$,

$$\left.\begin{aligned} e_{S1} &= E_A\varepsilon^{j(\omega t-\theta_f)} \\ i^{S1} &= I^A\varepsilon^{j(\omega t-\theta_f)} \end{aligned}\right\} \quad . \quad (135) \qquad \left.\begin{aligned} e_{S2} &= je_{S1} \\ i^{S2} &= ji^{S1} \end{aligned}\right\} \quad . \quad (136)$$

When the rotating reference frame has synchronous angular velocity, $p\theta_f = \omega$,

$$\left.\begin{aligned} e_{S1} &= E_A \\ i^{S1} &= I^A \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (137)$$

The balanced steady-state equation may be written down by putting $p = 0$ in eqns. (133) and using eqns. (136). Thus

$$i_{S1} = Ri^{S1} + jv\omega Li^{S1} \quad . \quad . \quad . \quad . \quad (138)$$

And at synchronous angular velocity $v = 1$, and

$$e_{S1} = Ri^{S1} + jXi^{S1} \quad . \quad . \quad . \quad . \quad (139)$$

The operator p has therefore the same significance in the transient eqns. (133) as in Park's eqns. (21).

(6.1) The 3-Phase Induction Motor

A 3-phase induction motor has been analysed by Stanley[9] by expressing the equations in terms of an equivalent 2-phase machine. The 3-phase rotor and stator currents, voltages and flux linkages are resolved along two axes, namely the axis of the stator phase A and the common axis of the stator phases B and C in quadrature with phase A. Since both rotor and stator quantities are resolved along the same two axes these are relatively stationary and fixed on the stator as in Fig. 2.

The equations for a balanced symmetrical motor with voltages applied to the stator may therefore be written

ds	e_{ds}
dr	0
qr	0
qs	e_{qs}

$=$

	ds	dr	qr	qs
ds	$R_1 + L_1p$	Mp		
dr	Mp	$R_2 + L_2p$	$L_2p\theta$	$Mp\theta$
qr	$-Mp\theta$	$-L_2p\theta$	$R_2 + L_2p$	Mp
qs			Mp	$R_1 + L_1p$

$\cdot$

ds	i^{ds}
dr	i^{dr}
qr	i^{qr}
qs	i^{qs}

(140)

or $e_m = R_{mk}i^k + L_{mk}pi^k + \Gamma_{kn,m}i^ki^n$.

where R_1 = Stator resistance per phase.
R_2 = Total rotor resistance per phase.
L_1 = Stator phase inductance.
L_2 = Rotor phase inductance.
M = Maximum value of mutual inductance stator/rotor phase.

In balanced steady state $i^{ds} = ji^{qs}$, $i^{dr} = ji^{qr}$ and p becomes $j\omega$, and $L_2(p - jp\theta)$ becomes $L_2(j\omega - jp\theta) = js\omega L_2$, where $s = 1 - v$ and $v = \dfrac{\text{angular velocity of rotor}}{\text{synchronous angular velocity}}$

Eqns. (140) therefore become, in balanced steady state,

$$e_{ds} = (R_1 + jX_1)i^{ds} + jX_mi^{dr} \quad . \quad . \quad . \quad (141)$$

$$0 = jsX_mi^{ds} + (R_2 + jsX_2)i^{dr} \quad . \quad . \quad . \quad (142)$$

When an induction motor is associated with a synchronous machine the motor equations may be written in terms of axes rotating with the flux as in Park's equations. In the balanced steady state, in this case, the vectors representing voltages, currents and flux linkages, are constant in magnitude and rotate synchronously in space around the stator of the machine. As shown in Fig. 3 there are two axes, S_3 and S_4, in quadrature on the stator, and two, S_1 and S_2, in quadrature on the rotor. The stator and rotor axes are both rotating synchronously and are again relatively stationary. One would therefore expect to find that the equations in this frame have a form similar to those of Stanley and Park. This is, in fact, the case. Both induced and generated voltages appear in the equations, but as in the synchronous-machine equations the steady-state equations of the motor are now obtained by putting such terms as Lp equal to zero, and only generated-voltage terms remain. It will be seen from the free-frame equations that, relative to these axes, the stator appears to rotate synchronously backwards and the rotor to rotate backwards at the angular velocity of slip.

The relationship among the currents in the stationary axes and those in the free frame may be written

$$\left.\begin{aligned} i^{ds} &= i^{S3}\cos\theta_1 - i^{S4}\sin\theta_1 \\ i^{dr} &= i^{S1}\cos\theta_1 - i^{S2}\sin\theta_1 \\ i^{qr} &= i^{S1}\sin\theta_1 + i^{S2}\cos\theta_1 \\ i^{qs} &= i^{S3}\sin\theta_1 + i^{S4}\cos\theta_1 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (143)$$

or $$i^m = C^m_\gamma i^\gamma \quad . \quad . \quad . \quad . \quad . \quad . \quad (144)$$

where $C^m_\gamma =$

$m \backslash \gamma$	S_3	S_1	S_2	S_4
ds	$\cos\theta_1$			$-\sin\theta_1$
dr		$\cos\theta_1$	$-\sin\theta_1$	
qr		$\sin\theta_1$	$\cos\theta_1$	
qs	$\sin\theta_1$			$\cos\theta_1$

(145)

The equations in the free frame are, as before,

$$e_\gamma = R_{\gamma\alpha}i^\alpha + L_{\gamma\alpha}pi^\alpha + 2[\alpha s,\gamma]i^\alpha p\theta + 2\Omega_{\gamma s,\alpha}i^\alpha p\theta \quad . \quad (146)$$

where $e_\gamma = C^m_\gamma e_m$

$R_{\gamma\alpha} = R_{mk}C^m_\gamma C^k_\alpha$

$L_{\gamma\alpha} = L_{mk}C^m_\gamma C^k_\alpha$

L_{mk} is shown in matrix 46.

It is found on carrying out the transformation that

$L_{\gamma\alpha} = L_{mk}$

$R_{\gamma\alpha} = R_{mk}$

therefore again $2[\alpha s,\gamma] = 0$

Now $$2\Omega_{\gamma s,\alpha} = 2\Omega_{ms,k}C^m_\gamma C^s_s C^k_\alpha + \frac{\partial C^\delta_m}{\partial\theta}C^m_\gamma L_{\delta\alpha} \quad . \quad . \quad (147)$$

and [eqn. (62)]

$2\Omega_{ms,k} =$

$m \backslash k$	ds	dr	qr	qs
ds				
dr			L_2	M
qr	$-M$	L_2		
qs				

. (148)

This term of the right-hand side of eqn. (147) is seen to transform tensorially (in this case to the same matrix)

$2\Omega_{ms,k}C^m_\gamma C^k_\alpha =$

$\gamma \backslash \alpha$	S_3	S_1	S_2	S_4
S_3				
S_1			L_2	M
S_2	$-M$	$-L_2$		
S_4				

$= 2\Omega'_{\gamma s,\alpha}$. (149)

This is the term referred to (by Kron) as the "torsion tensor," the name being taken from geometry.

Now $$\frac{\partial C^\delta_m}{\partial\theta}C^m_\gamma L_{\delta\alpha}p\theta = \frac{\partial C^\delta_m}{\partial\theta_1}C^m_\gamma L_{\delta\alpha}p\theta_1 \quad . \quad . \quad . \quad (150)$$

and

$\frac{\partial C^\delta_m}{\partial\theta_1}C^m_\gamma =$

$\gamma \backslash \delta$	S_3	S_1	S_2	S_4
S_3				-1
S_1			-1	
S_2		1		
S_4	1			

. (151)

Therefore

$\frac{\partial C^\delta_m}{\partial\theta_1}C^m_\gamma L_{\delta\alpha} =$

$\gamma \backslash \alpha$	S_3	S_1	S_2	S_4
S_3			$-M$	$-L_1$
S_1			$-L_2$	$-M$
S_2	M	L_2		
S_4	L_1	M		

. (152)

Thus $\left(2\Omega'_{\gamma s,\alpha}p\theta + \frac{\partial C^\delta_m}{\partial\theta_1}C^m_\gamma L_{\delta\alpha}p\theta_1\right) =$

$\gamma \backslash \alpha$	S_3	S_1	S_2	S_4
S_3			$-Mp\theta_1$	$-L_1p\theta_1$
S_1			$-L_2p\theta'$	$-Mp\theta'$
S_2	$Mp\theta'$	$L_2p\theta'$		
S_4	$L_1p\theta_1$	$Mp\theta_1$		

. (153)

where $p\theta'$ is the angular velocity of slip, namely $p\theta_1 - p\theta$ (which is equal to $-p\theta_f$). In the free frame the equations of the machine therefore become

			S_3	S_1	S_2	S_4		
S_3	e_{S3}	S_3	$R_1 + L_1p$	Mp	$-Mp\theta_1$	$-L_1p\theta_1$	S_3	i^{S3}
S_1	0	S_1	Mp	$R_2 + L_2p$	$-L_2p\theta'$	$-Mp\theta'$	S_1	i^{S1}
S_2	0	S_2	$Mp\theta'$	$L_2p\theta'$	$R_2 + L_2p$	Mp	S_2	i^{S2}
S_4	e_{S4}	S_4	$L_1p\theta_1$	$Mp\theta_1$	Mp	$R_1 + L_1p$	S_4	i^{S4}

(154)

Eqn. (154) may also be derived by starting from the holonomic machine in frame (i), Section 5, the relationship among the currents is then (Fig. 3)

$$\left.\begin{aligned} i^{ds} &= i^{S3}\cos\theta_1 - i^{S4}\sin\theta_1 \\ i^{a'} &= i^{S1}\cos\theta_f + i^{S2}\sin\theta_f \\ i^{b'} &= -i^{S1}\sin\theta_f + i^{S2}\cos\theta_f \\ i^{qs} &= i^{S3}\sin\theta_1 + i^{S4}\cos\theta_1 \end{aligned}\right\} \quad . \quad . \quad . \quad (155)$$

or $i^\alpha = C^a_\alpha i^\alpha$.

$C^a_\alpha =$

$a \backslash \alpha$	S_3	S_1	S_2	S_4
ds	$\cos\theta_1$			$-\sin\theta_1$
a'		$\cos\theta_f$	$\sin\theta_f$	
b'		$-\sin\theta_f$	$\cos\theta_f$	
qs	$\sin\theta_1$			$\cos\theta_1$

. (156)

The equations in the free frame, when transformed directly from the holonomic equations are [see Section 5, eqn. (78)]

$$e_\gamma = R_{\gamma\alpha}i^\alpha + L_{\gamma\alpha}pi^\alpha + 2[\alpha s,\gamma]i^\alpha i^s + 2\Omega_{\gamma s,\alpha}i^s i^\alpha$$

where in this case

$$2\Omega_{\gamma s,\alpha} = C^c_\gamma\frac{\partial C}{\partial\theta}L_{\delta\alpha} \quad . \quad . \quad . \quad . \quad (157)$$

This is the single non-holonomic object arising between the holonomic frame and the non-holonomic free frame.

Again $$L_{\gamma\alpha} = L_{ca}C^c_\gamma C^a_\alpha$$
$$R_{\gamma\alpha} = R_{ca}C^c_\gamma C^a_\alpha$$

For a symmetrical induction motor represented by Fig. 3, the inductance matrix 9 in Section 2 becomes

$L_{ca} =$

$c \backslash a$	ds	a'	b'	qs
ds	L_1	$M\cos\theta$	$-M\sin\theta$	
a'	$M\cos\theta$	L_2		$M\sin\theta$
b'	$-M\sin\theta$		L_2	$M\cos\theta$
qs		$M\sin\theta$	$M\cos\theta$	L_1

(158)

Using the relation $\theta = \theta_1 + \theta_f$ from Fig. 3, then

$L_{\gamma\alpha} =$

$\gamma \backslash \alpha$	S_3	S_1	S_2	S_4
S_3	L_1	M		
S_1	M	L_2		
S_2			L_2	M
S_4			M	L_1

. (159)

From matrix 156

$C^{c}_{\gamma}\frac{\partial C^{\delta}_{c}}{\partial\theta}=$

$\gamma \backslash \delta$	S_3	S_1	S_2	S_4
S_3				$-d\theta_1/d\theta$
S_1			$d\theta_f/d\theta$	
S_2		$-d\theta_f/d\theta$		
S_4	$d\theta_1/d\theta$			

(160)

And $C^{c}_{\gamma}\frac{\partial C^{\delta}_{c}}{\partial\theta}L_{\delta\alpha}\frac{d\theta}{dt}$ gives matrix 153.

It is seen that the induction-motor equations in a free frame rotating at uniform velocity, when derived from the stationary reference-frame equations, assume the form

$$e_\gamma = R_{\gamma\alpha}i^\alpha + L_{\gamma\alpha}pi^\alpha + \rho_1 L_{\gamma\alpha}i^\alpha p\theta + \rho_2 L_{\gamma\alpha}i^\alpha p\theta_1 \quad . \quad (161)$$

where $\rho_1 =$

	S_3	S_1	S_2	S_4
S_3				
S_1			-1	
S_2		1		
S_4				

. (162)

and $\rho_2 =$

	S_3	S_1	S_2	S_4
S_3				-1
S_1			-1	
S_2		1		
S_4	1			

. (163)

$\rho_1 L_{\gamma\alpha}$ is matrix 149 and $\rho L_{\gamma\alpha}$ is matrix 152.

Eqn. (161) may be written

$$e_\gamma = R_{\gamma\alpha}i^\alpha + L_{\gamma\alpha}pi^\alpha + G_{\gamma\alpha}i^\alpha p\theta + V_{\gamma\alpha}i^\alpha p\theta_1 \quad . \quad (164)$$

or in terms of flux vectors

$$\boldsymbol{e} = \boldsymbol{Ri} + p\boldsymbol{\psi} + \boldsymbol{B}p\theta + \boldsymbol{\varphi}p\theta_1 \quad . \quad . \quad . \quad (165)$$

Synchronous-machine equations in this form are discussed by Kron[15] and used in his hunting analysis.[20]

In balanced steady state

$$i^{S4} = ji^{S3} \qquad i^{S2} = ji^{S1} \quad . \quad . \quad . \quad . \quad (166)$$

The operator p is zero when this operates on the steady-state currents (but $p\theta_1 = v\omega$).

Thus, eqns. (154) become

$$\left.\begin{aligned} e_{S3} &= (R_1 + jv\omega L_1)i^{S3} + jv\omega Mi^{S1} \\ 0 &= js\omega Mi^{S3} + (R_2 + js\omega L_2)i^{S1} \end{aligned}\right\} \quad . \quad . \quad (167)$$

When the reference frame rotates at synchronous speed with respect to the stator winding, then $p\theta_1 = \omega$ and

$$\left.\begin{aligned} e_{S3} &= (R_1 + jX_1)i^{S3} + jX_m i^{S1} \\ 0 &= jsX_m i^{S3} + (R_2 + jsX_2)i^{S1} \end{aligned}\right\} \quad . \quad . \quad (168)$$

In this reference frame the operator p has the same significance as in the two-reaction theory of Park, and the free frame equations for a motor or network may be combined with those of synchronous machines when an interconnected network is being considered.

(6.2) Induction-Motor Torque Equations

The motor torque is given by eqn. (106).

Since $[\alpha\beta,s] = 0$, the equation (neglecting the mechanical friction term R_{st}) becomes:

$$\text{Impressed torque } f_s = L_{st}\frac{d^2\theta}{dt^2} + 2\Omega_{s\beta,\alpha}i^\beta i^\alpha \quad . \quad (169)$$

and at constant angular velocity

$$\text{Generated torque} = -2\Omega_{s\beta,\alpha}i^\beta i^\alpha$$

where, as in eqns. (62), $\Omega_{s\beta,\alpha}$ is the negative of $\Omega_{\gamma s,\alpha}$

Thus $\Omega_{s\beta,\alpha} = -C^{a}_{\alpha}\frac{\partial C^{\delta}_{a}}{\partial\theta}L_{\delta\beta} =$

$\alpha \backslash \beta$	S_3	S_1	S_2	S_4
S_3			$M\frac{d\theta_1}{d\theta}$	$L_1\frac{d\theta_1}{d\theta}$
S_1			$-L_2\frac{d\theta_f}{d\theta}$	$-M\frac{d\theta_f}{d\theta}$
S_2	$M\frac{d\theta_f}{d\theta}$	$L_2\frac{d\theta_f}{d\theta}$		
S_4	$-L_1\frac{d\theta_1}{d\theta}$	$-M\frac{d\theta_1}{d\theta}$		

. (170)

$i^\beta =$

β	
S_3	i^{S3}
S_1	i^{S1}
S_2	i^{S2}
S_4	i^{S4}

. . . . . (171)

$$\Omega_{s\beta,\alpha}i^\beta i^\alpha = i^{S1}i^{S4}M - i^{S2}i^{S3}M \quad . \quad . \quad . \quad (172)$$

or
$$\Omega_{s\beta,\alpha}i^\beta i^\alpha = \psi_d i^q - \psi_q i^d \quad . \quad . \quad . \quad . \quad (173)$$

where ψ_d and ψ_q are the flux linkages in axes S_1 and S_2, respectively.

The generated torque given by $\Omega_{s\beta,\alpha}i^\beta i^\alpha$ is seen to be that given by

$$f = G_{\gamma\alpha}i^\gamma i^\alpha \quad . \quad . \quad . \quad . \quad . \quad . \quad (174)$$

where
$$G_{\gamma\alpha} = \rho_1 L_{\gamma\alpha}$$

or[8]
$$f = i^* B \quad . \quad . \quad . \quad . \quad . \quad . \quad (175)$$

where i^* is the conjugate of i^γ.

(7) CONCLUSION

The analysis of electrical machines may often be simplified by transforming the variables and parameters from the real phase reference frame to real or fictitious stationary or rotating reference axes. Changes of reference frames similar to those demon-

strated in the paper have been used by Kron and Ku in the derivation of equivalent circuits for electrical machinery.[3,23] Tensor analysis is a most useful mathematical tool for handling such transformations. Tensor equations look complicated when first used because the notation is comparatively new to engineers. In practical application, however, when the equations of a machine are written in tensor form they are fairly simple and the technique of transformation and calculation of phenomena become a matter of routine procedure.

This method of handling machine problems ensures that the analysis is systematic and the equations are of a form that often leads to clearer concepts of the interactions of the various currents and fluxes in the system. It is possible to distinguish in the equations between terms that have existence in all reference frames and those which arise because of the reference frame chosen [compare for example the terms $G_{\gamma\alpha}i^{\alpha}p\theta$ and $V_{\gamma\alpha}i^{\alpha}p\theta_1$ in eqn. (164)].

It is intended that the foregoing presentation of Kron's work on the tensor equations of electrical machines should provide a groundwork on which may be built a more complete mathematical and physical analysis of machine stability problems by investigating the phenomena in various reference frames.

(8) ACKNOWLEDGMENTS

The author wishes to express his thanks to Prof. J. M. Meek of the University of Liverpool for his interest in the work, and to Dr. W. J. Gibbs of the British Thomson-Houston Co. Ltd., for discussions and advice during the preparation of the paper. He also wishes to express his appreciation of the encouragement he received in the earlier stages of the investigations from Prof. Emeritus F. J. Teago, formerly of the University of Liverpool, and Prof. P. P. Burns of the Queen's University of Belfast.

(9) REFERENCES

(1) KRON, G.: "Non-Riemannian Dynamics of Rotating Electrical Machinery," *Journal of Mathematics and Physics*, 1934, **13**, p. 103.

(2) KRON, G.: "The Application of Tensors to the Analysis of Rotating Electrical Machinery," *General Electric Review*, 1935 to 1938. Published in book form 1938 and 1942.

(3) KRON, G.: "Equivalent Circuits of Electric Machinery" (John Wiley, 1951).

(4) GIBBS, W. J.: "Limitations of Lagrangian Methods in Electrical Machine Theory," *Beama Journal*, 1950, **57**, pp. 243 and 382.

(5) GIBBS, W. J.: "Tensors in Electrical Machine Theory" (Chapman and Hall, 1952).

(6) KRON, G.: "Quasi-holonomic Dynamical Systems," *Journal of Applied Physics*, 1936, **7**, p. 143.

(7) SABBAGH, E. M.: "Application of 2-reaction Theory to Electric Motors," *Transactions of the American I.E.E.*, 1951, **70**, Part II, p. 1748.

(8) KRON, G.: "Short Course in Tensor Analysis" (John Wiley, 1942).

(9) STANLEY, H. C.: "An Analysis of the Induction Machine," *Transactions of the American I.E.E.*, 1938, **57**, p. 751.

(10) LYNN, J. W., and ALDRED, A. S.: "The Practical Application of Matrix Methods of Electrical Machine Analysis," *Beama Journal*, 1954, **61**, pp. 114 and 140.

(11) GIBBS, W. J.: "The Equations and Circle Diagram of the Schrage Motor," *Journal I.E.E.*, 1946, **93**, Part II, p. 621.

(12) CONCORDIA, C.: "Synchronous Machines" (John Wiley, 1951).

(13) WHITTAKER, E. T.: "Analytical Dynamics" (Cambridge University Press, 1937).

(14) SCHOUTEN, J., and STRUIK, D. J.: "Einfuhrung in die Neuren Methoden der Differentialgeometric," Band I (Springer, Berlin, 1935).

(15) KRON, G.: "Classification of the Reference Frames of a Synchronous Machine," *Transactions of the American I.E.E.*, 1950, **69**, Part II, p. 720.

(16) HOFFMANN, B.: "Kron's Non-Riemannian Electrodynamics," *Reviews of Modern Physics*, 1949, **21**, p. 535.

(17) WAGNER, C. F., and EVANS, R. D.: "Symmetrical Components" (McGraw-Hill Book Co., 1933).

(18) PARK, R. H.: "Two-Reaction Theory of Synchronous Machines," *Transactions of the American I.E.E.*, (i) 1929, **48**, p. 716, and (ii) 1933, **52**, p. 352.

(19) MCCONNELL, A. S.: "Applications of the Absolute Differential Calculus" (Blackie and Sons, London, 1931).

(20) KRON, G.: "A New Theory of Hunting," *Transactions of the American I.E.E.*, 1952, **71**, Part III, p. 859.

(21) HEFFRON, W. G., ROSENBERRY, G. M., and ROTHE, F. S.: "Generalized Hunting Equations of Power Systems," *ibid.*, p. 1095.

(22) CLARKE, E.: "Circuit Analysis of A.C. Power Systems" (John Wiley), Volumes I and II.

(23) KU, Y. H.: "Rotating-Field Theory and General Analysis of Synchronous and Induction Machines," *Proceedings I.E.E.*, Monograph No. 54 U, June, 1952 (**99**, Part IV, p. 410).

(24) KRON, G.: "Stationary Networks and Transmission Lines along Uniformly Rotating Reference Frames," *Transactions of the American I.E.E.*, 1949, **68**, Part II, p. 690.

(25) FORSYTH, A. R.: "A Treatise on Differential Equations" (Macmillan and Co., London, 1914. Fourth Edition), p. 309.

(10) APPENDIX

(10.1) Tensor Transformations[19]

Tensor analysis of electrical machines is largely concerned with transformations of machine equations.

Tensors are sets of quantities, often represented by matrices, which are (*a*) functions of a set of co-ordinates (variables) and (*b*) subject to certain conditions of transformation when the co-ordinates are changed. The basic laws of transformation are set out below.

Let P be a particular value given by n co-ordinates x^i $(i = 1, 2, \ldots n)$ in one reference system and by $\bar{x}^j (j = 1, 2, \ldots m)$ in another system. Let Q be a value close to P, given by $x^i + dx^i$ and $\bar{x}^j + d\bar{x}^j$. The two sets of differentials in the two co-ordinate systems are connected by the equations

$$d\bar{x}^j = \frac{\partial \bar{x}^j}{dx^i} dx^i \quad . \quad . \quad . \quad . \quad . \quad . \quad (176)$$

The infinitesimal displacement PQ gives an example of a "contravariant" transformation. The indices are written by convention, as superscripts.

Another form of transformation is given as follows. Consider a scalar A which is invariant in all co-ordinate systems. The partial derivatives of A with respect to the co-ordinates x^i in one reference frame are given by $A_i = \partial A / \partial x^i$. In another system with co-ordinates x^j the partial derivatives will be given by

$$\bar{A}_j = \frac{\partial A}{\partial x^i} \frac{\partial x^i}{\partial \bar{x}^j} = \frac{\partial \bar{A}}{\partial \bar{x}^j} \quad . \quad . \quad . \quad . \quad (177)$$

Consequently

$$\bar{A}_j = A_i \frac{\partial x^i}{\partial \bar{x}^j} \quad . \quad . \quad . \quad . \quad . \quad . \quad (178)$$

The vector whose components in the x's are partial derivatives A_i is the gradient of the scalar A (grad A). This is an example

of a "covariant" vector, or tensor of the first order: it has one index written as a subscript. In general terms the two different forms of transformation may be written[19]

First-order tensors (vectors):

$$\text{Contravariant} \quad \bar{u}^j = u^i \frac{\partial \bar{x}^j}{\partial x^i} \quad . \quad . \quad . \quad (179)$$

$$\text{Covariant} \quad \bar{v}_j = v_\pi \frac{\partial x^\pi}{\partial \bar{x}^j} \quad . \quad . \quad . \quad . \quad (180)$$

Second- and higher-order tensors:

$$\text{Contravariant} \quad \bar{u}^{ab} = u^{\alpha\beta} \frac{\partial \bar{x}^a}{\partial x^\alpha} \frac{\partial \bar{x}^b}{\partial x^\beta} \quad . \quad . \quad (181)$$

$$\bar{u}^{abc} = u^{\alpha\beta\gamma} \frac{\partial \bar{x}^a}{\partial x^\alpha} \frac{\partial \bar{x}^b}{\partial x^\beta} \frac{\partial \bar{x}^c}{\partial x^\gamma} \quad . \quad . \quad . \quad . \quad (182)$$

$$\text{Covariant} \quad \bar{v}_{jk} = v_{\pi\rho} \frac{\partial x^\pi}{\partial \bar{x}^j} \frac{\partial x^\rho}{\partial \bar{x}^k} \quad . \quad . \quad . \quad (183)$$

$$\bar{v}_{jkm} = v_{\pi\kappa\mu} \frac{\partial x^\pi}{\partial \bar{x}^j} \frac{\partial x^\kappa}{\partial \bar{x}^k} \frac{\partial x^\mu}{\partial \bar{x}^m} \quad . \quad . \quad . \quad . \quad (184)$$

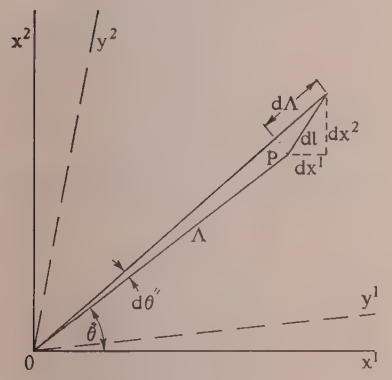


Fig. 6.—Cartesian, rectilinear and polar co-ordinates.

(10.1.1) Example of Simple Linear Transformation.

Let P be a point located by Cartesian co-ordinates x^1 and x^2 as in Fig. 6. A linear transformation to new rectilinear co-ordinates is given by the relation

$$\left.\begin{aligned} x^1 &= a_1^1 y^1 + a_2^1 y^2 \\ x^2 &= a_1^2 y^1 + a_2^2 y^2 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (185)$$

or $$x^\alpha = C^\alpha_{\cdot u} y^u \quad . \quad . \quad . \quad . \quad . \quad . \quad (186)$$

where $x^\alpha =$

α	
1	x^1
2	x^2

$\quad . \quad . \quad . \quad . \quad . \quad (187)$

$C^\alpha_{\cdot u} =$

α \ u	1	2
1	a_1^1	a_2^1
2	a_1^2	a_2^2

$\quad . \quad . \quad . \quad . \quad . \quad (188)$

$y^u =$

u	
1	y^1
2	y^2

$\quad . \quad . \quad . \quad . \quad . \quad (189)$

(The column index of matrices such as $C^\alpha_{\cdot u}$ is preceded by a dot.) In Cartesian co-ordinates the length of the vector OP is given by

$$(\Lambda)^2 = (x^1)^2 + (x^2)^2 \quad . \quad . \quad . \quad . \quad (190)$$

This length is invariant with respect to any change of co-ordinate system. In the new co-ordinate system, by substitution,

$$(\Lambda)^2 = y^1 y^1 (a_1^1 a_1^1 + a_1^2 a_1^2) + 2 y^1 y^2 (a_2^2 a_1^2 + a_2^1 a_1^1) + y^2 y^2 (a_2^1 a_2^1 + a_2^2 a_2^2) \quad . \quad (191)$$

or $$(\Lambda)^2 = g_{11} y^1 y^1 + g_{12} y^1 y^2 + g_{21} y^2 y^1 + g_{22} y^2 y^2 \quad . \quad (191a)$$

where $g_{11} = (a_1^1 a_1^1 + a_1^2 a_1^2)$, $g_{22} = (a_2^1 a_2^1 + a_2^2 a_2^2)$

and $g_{12} = g_{21} = (a_2^2 a_1^2 + a_2^1 a_1^1)$

In matrix notation $(\Lambda)^2 = g_{\alpha\beta} x^\alpha x^\beta \quad \left\{\begin{matrix} \alpha = 1, 2 \\ \beta = 1, 2 \end{matrix}\right\} \quad . \quad (192)$

In Cartesian co-ordinates

$g_{\alpha\beta} =$

α \ β	1	2
1	1	
2		1

$\quad . \quad . \quad . \quad . \quad . \quad (193)$

and $g_{\alpha\beta} x^\alpha x^\beta$ is

α	1	2
	x^1	x^2

$\cdot$

α \ β	1	2
1	1	
2		1

$\cdot$

β	
1	x^1
2	x^2

$\quad (194)$

The quantity $g_{\alpha\beta}$ is a covariant double tensor and transforms according to eqn. (183); thus from eqns. (183) and (188)

$$g_{uv} = g_{\alpha\beta} C^\alpha_{\cdot u} C^\beta_{\cdot v} \quad . \quad . \quad . \quad . \quad . \quad (195)$$

$C^\alpha_{\cdot u} =$

u \ α	1	2
1	a_1^1	a_1^2
2	a_2^1	a_2^2

$\quad . \quad . \quad . \quad . \quad . \quad (196)$

(the transpose of matrix 188)

and $g_{uv} =$

u \ α	1	2
1	a_1^1	a_1^2
2	a_2^1	a_2^2

$\cdot$

α \ β	1	2
1	1	
2		1

$\cdot$

β \ v	1	2
1	a_1^1	a_2^1
2	a_1^2	a_2^2

$\quad (197)$

or $g_{uv} =$

u \ v	1	2
1	$a_1^1 a_1^1 + a_1^2 a_1^2$	$a_2^2 a_1^2 + a_2^1 a_1^1$
2	$a_2^1 a_1^1 + a_2^2 a_1^2$	$a_2^1 a_2^1 + a_2^2 a_2^2$

$\quad . \quad . \quad . \quad (198)$

which gives eqns. (191) when u and v each take the values 1 and 2.

The tensor g_{uv} or $g_{\alpha\beta}$ is called the "fundamental tensor" or "metric tensor" of the system. The metric tensor is necessary in order to calculate invariant properties of vectors or tensors, for example, the length of vectors, the angle between vectors and parallel displacements.

(10.1.2) **Non-Linear Transformation.**

An example of a non-linear transformation of variables is given by the change from Cartesian to polar co-ordinates. In this case the transformation must be one between differentials of the respective co-ordinate systems. From Fig. 6 it is seen that

$$\left.\begin{aligned} x^1 &= \Lambda\cos\theta'' \\ x^2 &= \Lambda\sin\theta'' \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (199)$$

$$\left.\begin{aligned} dx^1 &= -\Lambda\sin\theta''d\theta'' + \cos\theta''d\Lambda \\ dx^2 &= \Lambda\cos\theta''d\theta'' + \sin\theta''d\Lambda \end{aligned}\right\} \quad . \quad . \quad (200)$$

Now $$dx^\alpha = C^\alpha_{\cdot k}dx^k \quad . \quad . \quad . \quad . \quad . \quad (201)$$

where $dx^\alpha =$

α	
1	dx^1
2	dx^2

$dx^k =$

k	
1	$d\Lambda$
2	$d\theta''$

and $C^\alpha_{\cdot k} =$

$\alpha \backslash k$	1	2
1	$\cos\theta''$	$-\Lambda\sin\theta''$
2	$\sin\theta''$	$\Lambda\cos\theta''$

. . . (202)

Calling the invariant line element of length of any vector, in Cartesian co-ordinates, dl,

$$(dl)^2 = (dx^1)^2 + (dx^2)^2 = g_{\alpha\beta}dx^\alpha dx^\beta \quad . \quad . \quad (203)$$

where $g_{\alpha\beta}$ is given in matrix 193.

In polar co-ordinates

$$(dl)^2 = g_{kn}dx^k dx^n \quad . \quad . \quad . \quad . \quad . \quad (204)$$

where $$g_{kn} = g_{\alpha\beta}C^\alpha_{\cdot k}C^\beta_{\cdot n} \quad . \quad . \quad . \quad . \quad . \quad (205)$$

thus $g_{kn} =$

$k \backslash n$	1	2
1	1	
2		$(\Lambda)^2$

. . . (206)

and $$g_{kn}dx^k dx^n = g_{11}d\Lambda d\Lambda + g_{22}d\theta''d\theta'' \quad . \quad . \quad (207)$$

$$g_{12} = g_{21} = 0$$

therefore

$$(dl)^2 = (d\Lambda)^2 + (\Lambda)^2(d\theta'')^2 = (dx^1)^2 + (dx^2)^2 \quad . \quad (208)$$

(10.2) Lagrange's Equation[8,13]

In dynamics the behaviour of a system may be calculated provided that the parameters of the system are known, i.e. the masses, inertias, etc., together with the forces and constraints acting on the system.

Certain properties of the system will be invariant under a transformation of co-ordinates, one of these invariants is the kinetic energy in the system.

Kinetic energy $T = \frac{1}{2}mv^2$.

Now $$(dl)^2 = g_{ab}dx^a dx^b$$

therefore $$\left(\frac{dl}{dt}\right)^2 = v^2 = g_{ab}\frac{dx^a}{dt}\frac{dx^b}{dt} = g_{ab}\dot{x}^a\dot{x}^b \quad . \quad . \quad (209)$$

Thus $$T = \tfrac{1}{2}mg_{ab}\dot{x}^a\dot{x}^b \quad . \quad . \quad . \quad . \quad . \quad (210)$$

Lagrange's dynamical equation in generalized co-ordinates x^c for a free system acted upon by forces f_c is written (neglecting potential energy in this case) thus:

$$\frac{d}{dt}\left(\frac{\partial T}{\partial\dot{x}^c}\right) - \frac{\partial T}{\partial x^c} = f_c$$

$$\begin{aligned} \frac{\partial T}{\partial\dot{x}^c} &= \tfrac{1}{2}m\frac{\partial(g_{ab}\dot{x}^a\dot{x}^b)}{\partial\dot{x}^c} \\ &= \tfrac{1}{2}mg_{ab}\dot{x}^b\frac{\partial\dot{x}^a}{\partial\dot{x}^c} + \tfrac{1}{2}mg_{ab}\dot{x}^a\frac{\partial\dot{x}^b}{\partial\dot{x}^c} \\ &= \tfrac{1}{2}mg_{ab}\dot{x}^b\delta^a_c + \tfrac{1}{2}mg_{ab}\dot{x}^a\delta^b_c \\ &= \tfrac{1}{2}mg_{cb}\dot{x}^b + \tfrac{1}{2}mg_{ac}\dot{x}^a \quad . \quad . \quad . \quad . \quad . \quad (211) \end{aligned}$$

where $\delta^a_c = 1, \ a = c$ $\qquad$ $\delta^b_c = 1, \ b = c$
$\qquad\quad = 0, \ a \neq c$ $\qquad\quad = 0, \ b \neq c$

$$\begin{aligned} \frac{d}{dt}\left(\frac{\partial T}{\partial\dot{x}^c}\right) &= \tfrac{1}{2}m\left(\frac{dg_{ab}}{dt}\dot{x}^b + \frac{dg_{ac}}{dt}\dot{x}^a + g_{cb}\ddot{x}^b + g_{ca}\ddot{x}^a\right) \\ &= \tfrac{1}{2}m\left(\frac{\partial g_{cb}}{\partial x^a}\frac{dx^a}{dt}\dot{x}^b + \frac{\partial g_{ac}}{\partial x^b}\frac{dx^b}{dt}\dot{x}^a\right) + \tfrac{1}{2}mg_{cb}\ddot{x}^b + \tfrac{1}{2}mg_{ca}\ddot{x}^a \\ &= \tfrac{1}{2}m\left(\frac{\partial g_{cb}}{\partial x^a} + \frac{\partial g_{ac}}{\partial x^b}\right)\dot{x}^a\dot{x}^b + mg_{cb}\ddot{x}^b \quad . \quad . \quad . \quad . \quad (212) \end{aligned}$$

$$\frac{\partial T}{\partial x^c} = \tfrac{1}{2}m\frac{\partial}{\partial x^c}(g_{ab}\dot{x}^a\dot{x}^b) = \tfrac{1}{2}m\frac{\partial g_{ab}}{\partial x^c}\dot{x}^a\dot{x}^b \quad . \quad . \quad (213)$$

Thus

$$\begin{aligned} \frac{d}{dt}\left(\frac{\partial T}{\partial\dot{x}^c}\right) - \frac{\partial T}{\partial x^c} &= \tfrac{1}{2}m\left(\frac{\partial g_{cb}}{\partial x^a} + \frac{\partial g_{ca}}{\partial x^b} - \frac{\partial g_{ab}}{\partial x^c}\right)\dot{x}^a\dot{x}^b + mg_{cb}\ddot{x}^b \\ &= m[ab,c]\dot{x}^a\dot{x}^b + mg_{cb}\ddot{x}^b \\ &= m\{[ab,c]\dot{x}^a\dot{x}^b + g_{cb}\ddot{x}^b\} \quad . \quad . \quad . \quad . \quad . \quad (214) \end{aligned}$$

The expression in compound brackets in eqn. (214) is that for the acceleration of a particle in terms of generalized variables or co-ordinates. The quantities mg_{ab} define the metric tensor L_{ab} in dynamics. This term comprises the moments of inertia of the system.

Lagrange's equation may therefore be written as in Section 2, eqn. (10):

$$f_c = L_{ca}\ddot{x}^a + [ab,c]\dot{x}^a\dot{x}^b \quad . \quad . \quad . \quad . \quad (215)$$

While neither of the terms on the right-hand side is a tensor by itself, the expression on the right-hand side as a whole is a tensor. This is illustrated by a transformation to new co-ordinates,

$$f_\gamma = C^c_\gamma f_c \quad . \quad . \quad . \quad . \quad . \quad . \quad (216)$$

and $$f_\gamma = L_{\gamma\alpha}\ddot{x}^\alpha + [\alpha\beta,\gamma]\dot{x}^\alpha\dot{x}^\beta \quad . \quad . \quad . \quad (217)$$

where $$[L_{\gamma\alpha}\ddot{x}^\alpha + (\alpha\beta,\gamma)\dot{x}^\alpha\dot{x}^\beta] = [L_{ca}\ddot{x}^a + (ab,c)\dot{x}^a\dot{x}^b]C^c \quad . \quad (218)$$

In an electrical network the equation corresponding to eqn. 215) becomes

$$e_w = L_{wv}\frac{d^2q^v}{dt^2} + [uv,w]\frac{dq^u}{dt}\frac{dq^v}{dt} \quad . \quad . \quad . \quad (219)$$

or

$$e_w = L_{wv}\frac{di^v}{dt} + [uv,w]i^u i^v \quad . \quad . \quad . \quad . \quad (220)$$

where the metric tensor L_{wv} comprises the self- and mutual nductances of the network elements. The voltage drop due to esistance may be added as an extra term $R_{wv}i^v$.

(10.3) Conditions necessary for an Equation to be Integrable[25]

If $$Adx + Bdy + Cdz = 0 \quad . \quad . \quad . \quad . \quad (221)$$

as an integral $$f(xyz) = K$$

which on integration gives

$$\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz = 0$$

hen $$\frac{\partial f}{\partial x} = aA, \quad \frac{\partial f}{\partial y} = aB, \text{ and } \frac{\partial f}{\partial z} = aC$$

Hence $$\frac{\partial}{\partial y}(aC) = \frac{\partial^2 f}{\partial y \partial z} = \frac{\partial^2 f}{\partial z \partial y} = \frac{\partial}{\partial z}(aB)$$

e. $$a\left(\frac{\partial B}{\partial z} - \frac{\partial C}{\partial y}\right) + B\frac{\partial a}{\partial z} - C\frac{\partial a}{\partial y} = 0$$

nd $$\frac{\partial}{\partial y}(aC) = \frac{\partial a}{\partial y}C + a\frac{\partial C}{\partial y} = B\frac{\partial a}{\partial z} + a\frac{\partial B}{\partial z}$$

Therefore $$a\left(\frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}\right) + B\frac{\partial a}{\partial z} - C\frac{\partial a}{\partial y} = 0 \quad . \quad . \quad [222(a)]$$

Similarly $$a\left(\frac{\partial C}{\partial x} - \frac{\partial A}{\partial z}\right) + C\frac{\partial a}{\partial x} - A\frac{\partial a}{\partial z} = 0 \quad . \quad . \quad [222(b)]$$

and $$a\left(\frac{\partial A}{\partial y} - \frac{\partial B}{\partial x}\right) + A\frac{\partial a}{\partial y} - B\frac{\partial a}{\partial x} = 0 \quad . \quad . \quad [222(c)]$$

Multiplying eqns. [222(*a*)], [222(*b*)] and [222(*c*)] by A, B and C, respectively, and then adding,

$$A\left(\frac{\partial B}{\partial z} - \frac{\partial C}{\partial y}\right) + B\left(\frac{\partial C}{\partial x} - \frac{\partial A}{\partial z}\right) + C\left(\frac{\partial A}{\partial y} - \frac{\partial B}{\partial x}\right) = 0 \quad . \quad (223)$$

If eqn. (221) is integrable this condition must be satisfied. Eqn. (221) may be written

$$A_1dx^1 + A_2dx^2 + A_3dx^3 = 0$$

and the condition for integrability becomes

$$A_1\left(\frac{\partial A_2}{\partial x^3} - \frac{\partial A_3}{\partial x^2}\right) + A_2\left(\frac{\partial A_3}{\partial x^1} - \frac{\partial A_1}{\partial x^3}\right) + A_3\left(\frac{\partial A_1}{\partial x^2} - \frac{\partial A_2}{\partial x^1}\right) = 0$$

. . . . . (224)

or in general $$a_{m,n}A_r + a_{n,r}A_m + a_{r,m}A_n = 0 \quad . \quad . \quad . \quad (225)$$

where $$a_{m,n} = \frac{\partial A_m}{\partial x^n} - \frac{\partial A_n}{\partial x^m} \quad . \quad . \quad . \quad . \quad (226)$$

If, therefore, a set of equations, such as

$$dx^a = C^a{}_m dx^m$$

is not integrable, then $$\frac{\partial C^h_a}{\partial x^b} \neq \frac{\partial C^h_b}{\partial x^a}$$

621.372.5

Monograph No. 118 R
Jan. 195

SOME FUNDAMENTAL PROPERTIES OF NETWORKS WITHOUT MUTUAL INDUCTANCE

By A. TALBOT, M.A., Ph.D.

(*The paper was first received* 19*th July, and in revised form* 24*th September*, 1954.
It was published as an INSTITUTION MONOGRAPH *in January*, 1955.)

SUMMARY

Elementary methods, avoiding the manipulation of determinants, are used to prove theorems, not previously enunciated, on the voltage and current gains obtainable from a resistance network, and to deduce therefrom some properties of the algebraic expressions for the voltages and currents in the network.

After a brief survey of known properties of general *LRC* networks, in terms of the complex frequency variable, λ, the above results are used to obtain properties of general networks without mutual inductance, and in particular a simpler formulation and proof of some properties of *RC* networks than were given in a recent paper by Fialkow and Gerst.[3]

Finally, attention is drawn to a neglected paper of Kirchhoff,[10] in which the foundations of the topology of networks were laid down, and use is made of Kirchhoff's results to obtain explicit expressions, not involving determinants, for the various network parameters.

LIST OF PRINCIPAL SYMBOLS

$\alpha, \beta, \gamma, \delta$ = Elements of chain matrix.
b = Number of branches.
D = Network determinant.
I or I_1 = Input current of a resistance network.
i_1, i_2 = Currents at sides 1, 2 of a general 4-terminal network.
I_h, I_k = Currents in branches R_h, R_k.
I_{pq} = Current in branch pq from p to q.
λ = Complex frequency variable $= \sigma + j\omega$.
m = Number of independent meshes $= b - n + 1$.
μ = Complex variable, such that $\lambda = \mu^2$.
n = Number of nodes.
N_h = Minors of D, to which the I_h are proportional.
$p, q, \ldots$ = Nodes of a resistance network.
R = Load resistance.
R_h, R_k = Branch resistances.
R_{pq} = Resistance of branch pq.
T = Voltage-gain, V_o/V_i.
$T_{1,2}(R)$ = Voltage-gain in load R when placed at side 2.
τ = Current gain, I_{pq}/I.
$\tau_{1,2}(R)$ = Current gain in load R when placed at side 2.
V_i, V_o = Input and output voltages of a resistance network.
v_1, v_2 = Voltages at sides 1, 2 of a general 4-terminal network.
V_{pq} = Potential fall from p to q.
y_{ij} = Elements of admittance matrix.
z_{ij} = Elements of impedance matrix.
$Z_h(\lambda)$ = Branch impedance in a general network.

(1) INTRODUCTION

In the last few years 4-terminal *RC* networks have come to be used more and more in place of *LC* (reactance) networks for transmission purposes. In a number of recent papers where various properties of *RC* networks are mentioned, the reader is generally referred for proofs to Cauer's classical paper,[1] or to Guillemin[2]—who himself quotes Cauer. Recently, Fialkow and Gerst[3] gave necessary and sufficient conditions for a rational function of the complex frequency variable (which we shall denote by λ) to be the transfer function of a 3- or 4-terminal *RC* network. The conditions relate to poles and zeros of the function, and to its values for real λ. To prove the necessity of the conditions, except those relating to the poles, the author resorted to a manipulation of determinants. Moreover, their enunciation of these conditions was unnecessarily complicated and it obscured the essential fact. For the conditions on the poles they referred to Cauer, who, however, did not prove them *a priori*, but himself quoted Routh[5] for results on the roots of $\det(R + \lambda S)$ and its principal minors, where R and S are matrices of positive-definite quadratic forms, namely in the present case the total dissipation of energy and twice the total electrostatic energy in the network (cf. Reference 4, p. 224).

The object of the present paper is to establish all the above-mentioned results, and others besides, by relatively simple means. For the most part they follow from theorems on resistance networks, which are intuitively obvious and must surely be known, but which the author has never seen explicitly stated. The reason why such theorems are applicable to general networks is simply that any *LRC* network without mutual inductance behaves like a resistance network for each real non-negative value of λ.

The remaining results are here deduced from well-known properties of reactance networks, including Foster's theorem. These are applicable since every *RC* network can be changed into an *LC* network with closely related network functions by replacing every resistor R by a coil of inductance equal to R.

(2) THE VOLTAGE-GAIN THEOREMS FOR RESISTANCE NETWORKS

We consider a network containing a finite number of nodes linked together in any way by branches of resistance R, where $0 \leqslant R \leqslant \infty$. The extreme values $R = 0$ (short-circuit branch) and $R = \infty$ (open-circuit) are allowed at present so that the most general results may be obtained for later application. Values of R other than zero and infinity will be described as finite values; zero and infinity will be called non-finite values. If any 2, 3 or 4 nodes are singled out, they become the terminals of a 2-, 3- or 4-terminal network.

We shall restrict discussion to connected networks, i.e. such that between any two nodes there is a path along branches of the network. Two nodes will be said to be directly linked if they are connected by a finite branch, or by several branches in parallel. If the network can be divided into two parts having just one node and no branches in common, the parts will be called complementary sub-networks, hinged at the common node.

If voltage or current is applied to such a network, currents will flow in the branches, and differences of potential will be established between the nodes. Let V_{pq} denote the fall in potential

Correspondence on Monographs is invited for consideration with a view to publication.
Dr. Talbot is in the Mathematics Department of Imperial College, London.

from any node p to any node q. Then if p, q, r are any three nodes,

$$V_{pr} = V_{pq} + V_{qr} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

If p and q are directly linked, by a branch of resistance R_{pq}, the current I_{pq} from p to q along R_{pq} is given both in sense and in magnitude by Ohm's law:

$$I_{pq} = V_{pq}/R_{pq} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

We take for granted Kirchhoff's voltage or mesh law [which is a corollary of eqns. (1) and (2)] and his current or node law. No appeal will be made to topological results, but a few simple arguments of a topological nature will be used.*

Let 0, 1, 2 be any three nodes of a resistance network (see Fig. 1). Taking 1 as the input terminal, 2 as the output terminal

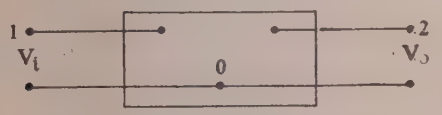


Fig. 1.—3-terminal resistance network.

and 0 as a joint input-and-output terminal, suppose that an input voltage $V_i = V_{10}$ is applied between 0 and 1, and let the resulting output voltage be $V_o = V_{20}$. Defining the voltage gain, T, as

$$T = \frac{V_o}{V_i} = \frac{V_{20}}{V_{10}}$$

our basic result is contained in the following theorem:

(2.1) **Theorem 1**

If T is the voltage-gain of a 3-terminal resistance network,

$$0 \leqslant T \leqslant 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The extreme values $T = 0$ and $T = 1$ can occur only if at least one branch resistance is non-finite, or if the output node belongs to a sub-network hinged at one of the input nodes and not containing the other.

Assume for definiteness $V_i > 0$, and suppose first that the network resistances are all finite.

If node 2 belongs to a sub-network, hinged at node 1 as in Fig. 2, $V_o = V_i$ and $T = 1$; if hinged at node 0, $T = 0$. (The

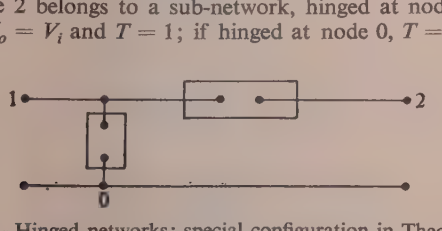


Fig. 2.—Hinged networks: special configuration in Theorem 1.

latter case is not illustrated.) If neither of these special cases occurs, suppose if possible that $T \geqslant 1$, and consider the node or nodes p for which V_{p0} has its maximum value in the network. (If maximum $V_{p0} = V_i$, include node 2 but exclude node 1.) Since the network is connected, and the special configurations are excluded, at least one of these nodes, say P, is directly linked to one or more of the remaining nodes q (including node 0 but excluding node 1), for which $V_{Pq} = V_{P0} - V_{q0} > 0$. Thus, by Ohm's law, all branch currents issuing from P are non-negative, and at least one is positive. Their sum is then positive, which contradicts Kirchhoff's current law. Hence $T < 1$.

Now, by solving the Kirchhoff equations, T can be found as a rational and therefore continuous function of the branch resistances. Since $T \leqslant 1$ when every resistance satisfies $0 < R < \infty$, this must hold also in the limiting case when some of the resistances have the value zero or infinity.

The remaining inequality, $T \geqslant 0$, may be proved in the same way, or else by noticing that (as pointed out by Fialkow and Gerst[3]) the complementary quantity $T' \equiv 1 - T$ is the gain, V_{21}/V_{10}, when 1 is taken as the joint terminal, so that $T' \leqslant 1$.

The limiting cases are illustrated by Fig. 3. Here, $T = 1$ if $R_{12} = 0$ or $R_{02} = \infty$, $T = 0$ if $R_{12} = \infty$ or $R_{02} = 0$.

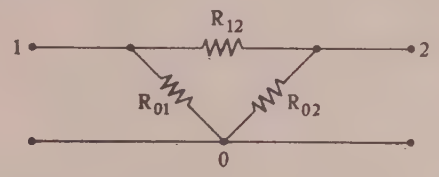


Fig. 3.—Network example for illustrating Theorem 1.

As regards a network of finite resistances without sub-networks hinged at either of the input nodes, Theorem 1 shows that the maximum and minimum potentials in the network can occur only at the input nodes. This may perhaps be compared with the fact that the maximum and minimum values of a non-constant function harmonic in a closed domain can occur only on the boundary.

Consider next the 4-terminal network shown in Fig. 4.

Fig. 4.—4-terminal resistance network.

Suppose V_i is applied to the input terminals 1′, 1 and V_o is the voltage produced at the output terminals 2′, 2. Clearly, $V_i = V_{1,1'}$, $V_o = V_{2,2'} = V_{2,1'} - V_{2',1'}$, so that the gain is

$$T = \frac{V_o}{V_i} = T_2 - T_{2'}$$

where T_2 and $T_{2'}$ are the gains in the 3-terminal networks having terminals 1′, 1, 2 and 1′, 1, 2′ respectively, with 1′ as joint terminal in each case. From Theorem 1 we deduce at once Theorem 2:

(2.2) **Theorem 2**

If T is the voltage-gain of a 4-terminal resistance network,

$$-1 \leqslant T \leqslant 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The extreme values $T = \pm 1$ can occur only if at least one branch resistance is non-finite, or if the two output nodes belong to separate sub-networks hinged to their respective complements at the two input nodes (as in Fig. 5, for example).

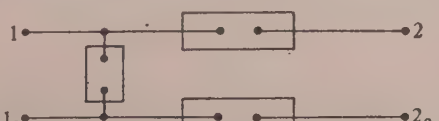


Fig. 5.—Hinged networks: special configuration in Theorem 2.

* The topology of networks has been fully discussed in a number of recent papers, e.g. by Ingram and Cramlet,[7] Synge,[8] Percival.[9] Actually, all the essential results, on trees, independent meshes, etc., were given by Kirchhoff[10] in an important but neglected paper, in which he showed how to write down the current in any branch of a resistance network by a topological inspection of the network, without using determinants. Kirchhoff's results are not used in the body of the present paper, but they are quoted in an Appendix, where, by their aid, it is shown how various network parameters can be similarly written down by inspection, without the necessity of solving equations.

It may be remarked that in both Theorems 1 and 2, if T has one of its extreme values because of the presence of non-finite resistances, the elimination of the latter by removing every infinite-resistance branch and merging the end-nodes of every zero-resistance branch will result in sub-networks of the appropriate kinds. For example, in Fig. 3, T is zero if R_{12} is infinite; and if R_{12} is removed, R_{02} forms a sub-network hinged to its complement, formed by R_{01}, at node 0.

(3) THE CURRENT-GAIN THEOREM FOR RESISTANCE NETWORKS

Suppose that an applied generator impresses current I (positive or negative) into node 1 and out of node 0 of a resistance network (see Fig. 6). By Ohm's law and Theorem 1 it is clear that, if

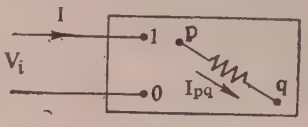


Fig. 6.—Resistance network.

$V_i > 0$, non-zero currents in branches ending at node 1 must in fact be all outward from the node, and at node 0 all inward. The sum in either case is equal to I. Thus if V_i is positive, so is I, which is obvious anyway since $V_i I$, the power dissipated, must be positive. (What is not obvious physically, without the help of Theorem 1, is that every single branch current at node 1 must be outward, and at node 0 inward.)

Next, let p be any other node at which there are non-zero currents. By Kirchhoff's current law, at least one of these must be inward to p, and at least one outward. Choose one of the branches carrying outward current, say the branch pp_1. The current in pp_1 is non-zero and inward to p_1. Let p_1p_2 be similarly one of the branches carrying current outward from p_1, and so on. Assuming the network resistances are all finite, we obtain in this way a sequence of nodes p, p_1, p_2, . . . with steadily decreasing potentials. Since the number of nodes in the network is finite by hypothesis, the sequence must ultimately, and in a finite number of steps, reach the node 0 which has the lowest potential. (If there is a sub-network hinged at node 0 and not containing node 1, all its nodes are equally at this minimum potential, but since there is zero current in all its branches, the sequence cannot reach any of its nodes.)

We have thus shown how to construct a path "following the current" from any node p, at which there are non-zero currents, to the node 0. Similarly one can always construct paths moving "against the current" from node p to node 1. Moreover, if pq is any branch of the network carrying positive current from p to q, we can always construct paths following the current from 1 to 0 via pq.

If the current in any branch pq is I_{pq}, positive or negative, from p to q, while I is the applied current, the current-gain τ in pq is defined as the ratio I_{pq}/I.

With the above preliminaries, it is easy to prove Theorem 3:

(3.1) Theorem 3

If τ is the current gain in any directed branch of a resistance network,

$$-1 \leqslant \tau \leqslant 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

and in the 3-terminal case when the branch begins at the positive input node or ends at the negative input node,

$$0 \leqslant \tau \leqslant 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5a)$$

The extreme values, ± 1 in eqn. (5) or 0, 1 in eqn. (5a), can occur only if at least one branch resistance is non-finite, or if the branch considered is the sole link between two sub-networks to which the two input nodes respectively belong (or, in the 3-terminal case, between one input node and a sub-network containing the other).

Suppose first the network resistances are all finite. Taking the general, 4-terminal, case, let pq be the branch considered, with current I_{pq}, positive or negative, flowing from p to q. Assume for definiteness that I, taken positively, flows as in Fig. 6, i.e. $V_i > 0$, and that $I_{pq} > 0$. Consider all possible paths "following the current" from q to 0, and let S be the set of all nodes included in these paths. There are now two cases: either all inward currents, other than I_{pq}, reaching the nodes of S arrive from other nodes of S, or else this is not true. In the first case, every inward current at every node of S except q is an outward current at some other node of S. Consequently, if we express Kirchhoff's current law for each node of S in the form

(sum of inward currents) = (sum of outward currents)

and add all the equations, all inter-node currents cancel out and we find that I_{pq} is equal to I. Since, by definition of S, all outward currents from nodes of S (except node 0) lead to other nodes of S, while in the case considered no inward currents reach nodes of S (other than q) except from other nodes of S, it is clear that the nodes of S and their interconnecting branches form a sub-network containing node 0 and hinged at q.

In the second case, there is a surplus of uncancelled inward currents when the node equations are added, whence $I_{pq} < I$. In this case there must be nodes of S other than q linked with nodes outside S.

By considering similarly paths "against the current" from p to 1 we find that $I_{pq} = I$ if there is a sub-network containing node 1 and hinged at p; otherwise $I_{pq} < I$.

The theorem thus follows if all network resistances are finite. If any are non-finite, the result $I_{pq} \leqslant I$ must still hold, by continuity.

The proof for the 3-terminal case is much simpler and will be omitted. The network in Fig. 7(*a*) illustrates the situation when

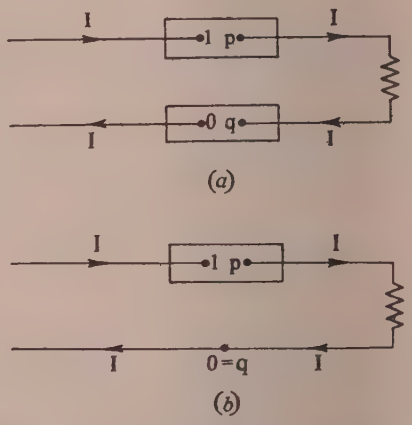


Fig. 7.—Special configurations in Theorem 3.
(*a*) 4-terminal case, (*b*) 3-terminal case.

$I_{pq} = I$ in the 4-terminal case if all network resistances are finite, or after any non-finite resistances have been eliminated. Fig. 7(*b*) illustrates the same situation in the 3-terminal case.

(4) ALGEBRAIC CONSEQUENCES OF THE GAIN THEOREMS

Consider a resistance network as in Fig. 6, having n nodes and b branches (assumed now to be all of finite resistance), including the generator branch between 0 and 1. Applying the current law to each node we obtain n equations involving only the branch currents, say I_k, in arbitrarily assigned directions. Of these equations only $n - 1$ can be independent, for the sum

of all the equations gives an identity, since every current is inward at one node and outward at another. Any $n-1$ of the equations are in fact independent, for any set of ρ nodes, $\rho \leqslant n-1$, has at least one node, say p, linked to a node, say q, outside the set: I_{pq} then occurs in only one of the ρ node equations, so that these cannot all be explicitly involved in a linear relation.

Next, Kirchhoff's voltage law gives an equation of the form

$$\Sigma \pm R_k I_k = V_i \text{ or } 0$$

for each mesh (closed circuit) that can be traced in the network, the right-hand side being V_i if the generator occurs in the mesh, otherwise 0. Suppose that the maximum number of independent equations, i.e. the rank of the set, is m. We then have altogether $m+n-1$ independent equations, not all homogeneous, for the b branch currents, so that $m+n-1 \leqslant b$. If $m+n-1 < b$, there would be solutions in terms of one or more parameters. In this case we could, by suitable choice of the parameters, either reverse the sign of the input current, I, which as we have seen is impossible for physical networks, i.e. those containing positive resistances; or if I should happen to be independent of the parameters, make one of the branch currents arbitrarily large, in violation of Theorem 3. In other words, for positive resistances the equations must have a unique solution, and $m = b-n+1$.*

On solving the equations we obtain

$$\frac{I_h}{V_i} = \frac{N_h}{D} \quad (h = 1, 2, \ldots, b) \quad . \quad . \quad . \quad . \quad (6)$$

where D, the network determinant, is of order b, the elements in its kth column being $\pm R_k$, ± 1 or 0, and N_h is a sum of determinants of order $b-1$, involving all columns of D except the hth. Thus N_h does not involve R_h. Examination of the forms of D and the N_h shows they are homogeneous polynomials in the R_k of degrees m and $m-1$, linear in each individual R_k, i.e. they are sums of multiples of products of m or $m-1$ different R_k values.

If we take the input current I as I_1, say, the input resistance of the network, between nodes 0 and 1, is

$$R = \frac{V_i}{I} = \frac{D}{N_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

By Theorem 2, every $|R_k I_k|$ is equal to or less than V_i. Thus if V_i is finite, so is every I_k, and hence also the input current I, which is a sum of branch currents. Conversely, if I is finite, so is every I_k, by Theorem 3, and hence also V_i, which can be expressed in the form $\Sigma \pm R_k I_k$.

Assuming, then, positive resistances, and finite V_i and I, the currents I_k must be unique and finite, and therefore D cannot vanish. It follows that all terms in D must have the same sign, for otherwise D, being linear in each R_k, could be made to vanish for suitable positive R_k values. Since the sign of D changes if any two of the equations are interchanged, we may assume all terms of D have positive signs.†

If p and q are any two nodes, there is a path from p to q along branches of the network, and by eqns. (1), (2) and (6),

$$\frac{V_{pq}}{V_i} = \Sigma \pm \frac{R_k N_k}{D}$$

summation being along the path, and the signs being positive or negative according to whether the pre-assigned sense of I_k follows or opposes the path. In particular we may write the voltage-gain in Theorem 1 as

$$T = \frac{V_{20}}{V_i} = \frac{C}{D}$$

where C, like D, is homogeneous of degree m in the R_k and linear in each R_k. We shall now show that every term in C is a term of D, with the same coefficient.

First, suppose the term $R_{a(1)}R_{a(2)}\ldots R_{a(m)}$ appears in C but not in D. By making all the $R_{a(s)}$ ($s = 1, \ldots m$) large compared with the remaining R_k, we can make this term large compared with every term of D, since no such term can have more than $m-1$ of the factors $R_{a(s)}$. Thus we can make T indefinitely large, which is impossible by Theorem 1. Next, suppose D has the term $aR_{a(1)}\ldots R_{a(m)}$ ($a > 0$) and C has $bR_{a(1)}\ldots R_{a(m)}$. On making the $R_{a(s)}$ infinitely large, T tends to b/a. Now if any R_k tends to infinity, I_k tends to zero since $|R_k I_k| \leqslant V_i$, by Theorem 2, so that the removal of R_k has no effect on the network. Consider the network remaining after the $R_{a(s)}$ are removed. This has $T = b/a$ for all values of the branch resistances. Now there must be a path from node 2 to node 0 or 1: otherwise T would be indeterminate. If there is a path to node 0, by making all its resistances zero we can make $T = 0$; similarly, if there is a path to node 1, we can make $T = 1$. It follows that $b = 0$ or $b = a$.

Thus finally we may write

$$T = \frac{D'}{D} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8a)$$

where D' comprises some of the terms of D.

Since T for a 4-terminal network is the difference between the gains of two 3-terminal networks, we have for a 4-terminal network

$$T = \frac{D' - D''}{D} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8b)$$

where D' and D'' each comprise some of the terms of D.

Again, N_1 cannot vanish for positive R_k values, and must be a sum of positive terms, since $I_1 > 0$ when $V_i > 0$. If the branch considered in Theorem 3 is R_h, then

$$\tau = \frac{I_h}{I_1} = \frac{N_h}{N_1} = \frac{N' - N''}{N_1}, \text{ say} \quad . \quad . \quad . \quad (9a)$$

where N' and $-N''$ are the sums of the positive and negative terms in N_h, and by Theorem 3 it is easy to prove that N' and N'' both comprise some of the terms of N_1. (Starting in the same way as above, we obtain a reduced network having $\tau = b/a$ for all resistance values. If b is not zero, there is a path following the current between the input nodes, via pq. Making all resistances infinite except those on this path gives $\tau = \pm 1$, i.e. $b = \pm a$.) Moreover, if R_h ends at one of the input nodes (the 3-terminal case), and the direction of I_h is appropriately prescribed, then

$$\tau = \frac{N'}{N_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9b)$$

where N' comprises some of the terms of N_1.

It is of interest to note that all the results of Sections 2 and 3 continue to hold even if the network resistors are non-linear, i.e. if the voltage drop V_k in the branch k is an arbitrary function of the current I_k flowing through it, provided that the "resistances" V_k/I_k are always positive. This is true even if the "slope resistances" dV_k/dI_k are sometimes negative; but if the latter is the case, then the uniqueness, demonstrated in the present Section, of the current system produced by a given

* This holds, of course, whatever the values of the resistances, and is in fact a form of Euler's topological formula.

† Consideration of the form of D shows that the coefficients must be integers. In fact[10] they are all equal to unity (see the Appendix). We do not, however, need this result.

applied voltage, need no longer hold. For example, if an e.m.f. of 6 units is applied to two resistors in series, having respectively the voltage laws $V = I$ and $V = 10I/(I^2 + 1)$, then the resulting current I can be equal to 1, 2, or 3 units.

However, if dV_k/dI_k is positive for every resistor, then it is not difficult to show that the current system is necessarily unique. For suppose that with a given applied voltage (or even a given set of applied voltages and currents) two different systems of network currents and voltages can result. Let the differences between these, in any branch k, be represented by i_k and v_k respectively, and write $r_k = v_k/i_k$. Then for each branch, $r_k > 0$, since $dV_k/dI_k > 0$. If now the set of node and mesh equations satisfied by one system is subtracted from that satisfied by the other, all the applied quantities will cancel, and the resulting set of equations will be the same as for a system of currents i_k flowing in a network of positive resistances r_k, with no applied voltage or current. The determinant of the set must be positive, as proved earlier in this Section, and the equations cannot be satisfied by currents i_k not all zero.

(5) PROPERTIES OF GENERAL NETWORKS

We now consider networks containing not only resistance but also capacitance and inductance. There are three possible ways of analysing such networks, namely (*a*) application of Kirchhoff's laws as in Section 4, (*b*) node analysis, i.e. use of the current law together with Ohm's law, and (*c*) mesh analysis, i.e. application of the voltage law to independent meshes in which fictitious circulating currents are assumed to flow. Method (*a*) has the disadvantage for theoretical purposes that it does not readily yield the reciprocity theorem or an equivalent, which is an essential result for passive networks, (*b*) may be awkward to apply if mutual inductances are present, while (*c*) requires a topological consideration of the network. For our present purpose, method (*c*) is the most convenient.

Consider, then, a general 4-terminal network as in Fig. 8.

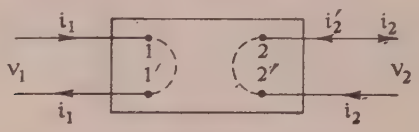


Fig. 8.—General 4-terminal network.

Select m independent meshes, including incomplete meshes through the input and output nodes, as indicated, and let circulating currents $i_1, i_2 = -i'_2, i_3, \ldots i_m$ flow in them. Assuming that all currents and voltages in the system are proportional to $\varepsilon^{\lambda t}$, where λ is the complex frequency variable $\sigma + j\omega$ (i.e. assuming steady-state working), the mesh equations may be written thus:

$$\left.\begin{aligned} \sum_1^m Z_{1s}(\lambda) i_s &= v_1 \\ \sum_1^m Z_{2s}(\lambda) i_s &= -v_2 \\ \sum_1^m Z_{rs}(\lambda) i_s &= 0, \quad r = 3, \ldots m \end{aligned}\right\} \quad . \quad . \quad . \quad (10)$$

Here, each $Z_{rs}(\lambda)$ is of the form $L_{rs}\lambda + R_{rs} + 1/C_{rs}\lambda$, and takes into account elements common to meshes r and s, and also, through L_{rs}, any mutual inductance between coils of r and s. As is easily seen, the matrix $[Z_{rs}]$ is symmetrical. If we wish to consider a network with only two terminals, say 1′, 1, we omit mesh 2 and the second of the above equations.

Solving the equations, we obtain, in the 2-terminal case,

$$i_1 = Y(\lambda) v_1 = \frac{1}{Z(\lambda)} v_1 \quad . \quad . \quad . \quad . \quad . \quad (11)$$

and in the 4-terminal case, dropping the argument λ for brevity,

$$\left.\begin{aligned} i_1 &= y_{11} v_1 + y_{12} v_2 \\ -i_2 = i'_2 &= y_{21} v_1 + y_{22} v_2 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (12)$$

Here, since $[Z_{rs}]$ is symmetrical,

$$y_{12} = y_{21} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

which is equivalent to a form of the reciprocity theorem.

Solving eqns. (12) for v_1, v_2, and noting eqn. (13), we obtain

$$\left.\begin{aligned} v_1 &= z_{11} i_1 + z_{12} i'_2 \\ v_2 &= z_{12} i_1 + z_{22} i'_2 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (14)$$

where, among other relations between y_{11}, etc., and z_{11}, etc., we note that

$$\frac{1}{y_{11}} = z_{11} - \frac{z_{12}^2}{z_{22}} \quad . \quad . \quad . \quad . \quad . \quad (15)$$

Alternatively, we may solve eqns. (12) for v_1, i_1:

$$\left.\begin{aligned} v_1 &= \alpha v_2 + \beta i_2 \\ i_1 &= \gamma v_2 + \delta i_2 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (16)$$

Here*

$$\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = -\frac{1}{y_{12}} \begin{bmatrix} y_{22} & 1 \\ y_{11}y_{22} - y_{12}^2 & y_{11} \end{bmatrix} = \frac{1}{z_{12}} \begin{bmatrix} z_{11} & z_{11}z_{22} - z_{12}^2 \\ 1 & z_{22} \end{bmatrix} \quad (17)$$

Obviously, $$\alpha\delta - \beta\gamma \equiv 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

which is a consequence of eqn. (13) and thus also equivalent to the reciprocity theorem.

All the functions of λ introduced in these equations are clearly rational and real.

In Fig. 8, if a resistance R is connected across the terminals 2′, 2, $v_2 = Ri_2$ and, by eqns. (16), the gains from side 1 to side 2 are, with an obvious notation,

$$\text{voltage gain: } T_{1,2}(R) = \frac{v_2}{v_1} = 1\Big/\left(\alpha + \frac{\beta}{R}\right) \quad . \quad . \quad (19)$$

$$\text{current gain: } \tau_{1,2}(R) = \frac{i_2}{i_1} = 1/(\delta + R\gamma) \quad . \quad . \quad (20)$$

Similarly, if R is connected across 1′, 1, the gains from side 2 to side 1 are

$$T_{2,1}(R) = 1\Big/\left(\delta + \frac{\beta}{R}\right), \quad \tau_{2,1}(R) = 1/(\alpha + R\gamma)$$

We may note that the voltage gain T at any pair of nodes of a network becomes $T_{1,2}(\infty)$, i.e. $1/\alpha$, on taking the terminals 2′, 2 at the nodes, while the current gain τ in any branch becomes $\tau_{1,2}(0)$, i.e. $1/\delta$, on taking 2′, 2 both at one end of the branch.

In the 2-terminal case [see eqn. (11)], the ratio of input voltage to input current is the input impedance function $Z(\lambda)$, and its reciprocal is the admittance function $Y(\lambda)$. $Z(\lambda)$ is a positive function, i.e.

$$RZ(\lambda) \geqslant 0 \text{ for } R\lambda \geqslant 0 \quad . \quad . \quad . \quad . \quad . \quad (21)$$

as was first shown by Cauer,[1] by energy considerations. Rational functions such as $Z(\lambda)$ which are real and positive were called

* This notation is similar to that normally used for these network parameters, and differs from that adopted for special reasons by the author in a previous paper.[11]

positive-real by Brune.[12] Clearly, if Z is positive-real, so is $Y \equiv 1/Z$. Also, it follows from the definition that

$$Z(\lambda) \geqslant 0 \text{ if } \lambda \geqslant 0 \quad . \quad . \quad . \quad . \quad . \quad (22)$$

In Fig. 9, the input impedances at 1′, 1 when terminals 2′, 2 are left open or are short-circuited, are z_{11} and $1/y_{11}$ respectively.

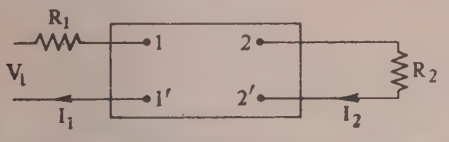


Fig. 9.—4-terminal network under working conditions.

Thus y_{11} and z_{11}, and similarly y_{22} and z_{22}, are positive-real. As a particular consequence,

$$y_{11}, y_{22}, z_{11}, z_{22} \geqslant 0 \text{ for } \lambda \geqslant 0 \quad . \quad . \quad . \quad (23)$$

Using eqn. (21), Brune showed that if $Z(\lambda)$ is positive-real it cannot have poles or zeros in the half-plane $R\lambda > 0$, and that at conjugate imaginary poles, say $\pm j\omega_0$, the residues are positive and equal, so that Z has a real partial-fraction term of the form

$$\frac{k}{\lambda - j\omega_0} + \frac{k}{\lambda + j\omega_0} = \frac{2k\lambda}{\lambda^2 + \omega_0^2} (k > 0) \quad . \quad . \quad (24)$$

(6) REACTANCE NETWORKS

A reactance network is one composed only of coils and condensers. Every Z_{rs} in eqns. (10) is then an odd function, and so are Y and Z in eqn. (11). Such functions, i.e. positive-real and odd, are called reactance functions. All their poles and zeros must lie on the imaginary axis, for if λ_0, in the left half-plane, is a pole or zero, so is $-\lambda_0$, in the forbidden right half-plane. Thus, if $Z(\lambda)$ is a reactance function, it must have an expansion of the form

$$Z(\lambda) = \frac{k_0}{\lambda} + k_\infty \lambda + \sum_r \frac{2k_r\lambda}{\lambda^2 + \omega_r^2} \ (k\text{'s} \geqslant 0) \quad . \quad (25)$$

Z has a zero at the origin if $k_0 = 0$, otherwise a pole; and a zero at infinity if $k_\infty = 0$, otherwise a pole. The residue at $\pm j\omega_r$ is k_r, and at 0 is k_0. For the sake of uniformity k_∞ will be called the residue at infinity. (This differs from the usual definition in function theory.)

Clearly $Z(j\omega)$ is imaginary, say $jX(\omega)$, where

$$X(\omega) = k_\infty \omega - \frac{k_0}{\omega} - \sum_r \frac{2k_r\omega}{\omega^2 - \omega_r^2}$$

Obviously $dX/d\omega > 0$* and it is easy to deduce Foster's theorem,[6] the essence of which is that the poles and zeros of $Z(\lambda)$ alternate on the imaginary axis.

Coming next to a 4-terminal reactance network, it is clear that y_{11}, y_{22}, z_{11}, z_{22} are all reactance functions; y_{12} and z_{12} are odd, but not necessarily reactances. However, by eqn. (15) we have

$$z_{12}^2 = z_{11}z_{22} - Zz_{22}$$

where $Z = 1/y_{11}$. All functions on the right are reactances and have simple poles on the imaginary axis. These include all the poles of z_{12}, which must therefore also be simple, and if the residues of z_{11}, z_{12}, z_{22}, Z at a pole of any or all of them are k_{11}, k_{12}, k_{22}, K, then

$$k_{11}k_{22} - k_{12}^2 = Kk_{22} \quad . \quad . \quad . \quad . \quad . \quad (26)$$

* This follows also directly from (21).

i.e. the determinant of the residues of the z's is non-negative. It follows that z_{12} can have a pole only where z_{11} and z_{22} have a common pole. Similar results hold for the y's. One consequence is that α, i.e. z_{11}/z_{12} or $-y_{22}/y_{12}$, and δ, i.e. z_{22}/z_{12} or $-y_{11}/y_{12}$ can have zeros only at zeros of their numerators. Their poles, on the other hand, may be either numerator poles or denominator zeros.

(7) PROPERTIES OF GENERAL NETWORKS WITHOUT MUTUAL INDUCTANCE

The impedance of any branch of a network is of the form $Z_h = L\lambda + R + 1/C\lambda$, where L, R and C are non-negative. For any real non-negative λ, Z_h is real and non-negative, and it is clear that if there is no mutual inductance to upset the relation $V_h = Z_hI_h$ (generalized Ohm's law) for each branch, then for each real non-negative λ the network is equivalent to a physical resistance network with branches of resistance $Z_h(\lambda)$. The gain theorems may be applied to eqns. (19) and (20), where R may have any non-negative value, and we may therefore state, in addition to eqn. (23), which holds for all passive networks:

For a 3-terminal network without mutual inductance, if $\lambda \geqslant 0$, $0 \leqslant T_{1,2}(R), \tau_{1,2}(R) \leqslant 1$ for all values of R equal to or greater than zero, whence

$$\beta, \gamma \geqslant 0, \quad 1 \leqslant \alpha, \delta \leqslant \infty \quad . \quad . \quad . \quad . \quad (27)$$

In terms of the y's and z's, these give:

$$0 \leqslant -y_{12} \leqslant y_{11}, y_{22}, \quad 0 \leqslant z_{12} \leqslant z_{11}, z_{22} \quad . \quad . \quad (28)$$

For a 4-terminal network without mutual inductance, if $\lambda \geqslant 0$, $-1 \leqslant T_{1,2}(R), \tau_{1,2}(R) \leqslant 1$ for all values of R equal to or greater than zero, whence

$$\beta/\alpha, \gamma/\delta \geqslant 0, \quad 1 \leqslant |\alpha|, |\delta| \quad . \quad . \quad . \quad . \quad (29)$$

(while $\alpha/\delta = y_{22}/y_{11} \geqslant 0$ by eqn. (23), so that α, β, γ, δ must be all non-negative or all non-positive). In terms of the y's and z's,

$$\left.\begin{aligned} -y_{11}, -y_{22} \leqslant y_{12} \leqslant y_{11}, y_{22} \\ -z_{11}, -z_{22} \leqslant z_{12} \leqslant z_{11}, z_{22} \end{aligned}\right\} \quad . \quad . \quad . \quad (30)$$

The same overall results are obtained by using $T_{2,1}$ and $\tau_{2,1}$ instead of $T_{1,2}$ and $\tau_{1,2}$.

If a network is "crossed," i.e. if the two input or output nodes are interchanged, then α, β, γ, δ all change sign. If a 3-terminal network is crossed it becomes effectively a 4-terminal network.

In all cases above, if we assume that none of the network branches is identically a short-circuit or an open-circuit, the extreme values indicated can be taken only for $\lambda = 0$ and $\lambda = \infty$ (the only non-negative values of λ for which any allowed Z_h can be equal to zero or infinity), unless one of the special configurations mentioned in the three gain theorems occurs.

Again, all the results of Section 4 (except for the last two paragraphs relating to non-linear networks) may be applied, with every R_h replaced by $Z_h(\lambda)$. Polynomials in the R's become polynomials in λ, divided by a power of λ. All coefficients in every term of D and N_1 are positive, and we can deduce at once, from eqns. 8(a) to 9(b):

For a 3-terminal network without mutual inductance, the voltage gain T can be expressed in the form

$$T = \frac{a_0\lambda^n + a_1\lambda^{n-1} + \ldots + a_n}{c_0\lambda^n + c_1\lambda^{n-1} + \ldots + c_n}$$

where $$0 \leqslant a_i \leqslant c_i \ (i = 0, 1, \ldots n)\dagger \quad . \quad . \quad . \quad (31)$$

The current gain τ in any network branch ending at one of the

† If the numerator and denominator have common factors, this form may not persist after their cancellation.

input nodes has the same form, and the inequalities in eqn. (31) again hold (if the branch-current direction is suitably prescribed).

For a 4-terminal network without mutual inductance, the voltage gain can be expressed in the form

$$T = \frac{b_0\lambda^n + b_1\lambda^{n-1} + \ldots + b_n}{c_0\lambda^n + c_1\lambda^{n-1} + \ldots + c_n}$$

where $$-c_i \leqslant b_i \leqslant c_i \ (i = 0, 1, \ldots n)\dagger \quad . \quad . \quad (32)$$

The same holds for the current gain in any network branch.

It may be remarked that if four numbers α, β, γ and δ, all of the same sign, are given, satisfying eqn. (18), and such that $|\alpha|$ and $|\delta|$ are equal to or greater than unity, then it is possible to construct a 3- or 4-terminal resistance network having these as parameters: a T-network, having horizontal arms $(|\alpha| - 1)/|\gamma|$, $(|\delta| - 1)/|\gamma|$ and vertical arm $1/|\gamma|$, and crossed if α, etc., are all negative, may, for example, be used. It is therefore unlikely that further properties of general networks without mutual inductance can be deduced by the method of this Section, i.e. comparison with resistance networks.

(8) PROPERTIES OF *RC* NETWORKS

All the results of Section 7 hold, in particular, for all *RC* networks. Additional results are obtained by noting that the impedance of a typical branch containing R and C in series is $Z_h(\lambda) = R + 1/\lambda C$, while the impedance of the same branch when the resistor is replaced by a coil of inductance equal to R is $Z_h^*(\lambda) = R\lambda + 1/\lambda C = \lambda Z_h(\lambda^2)$. It follows that in any *RC* network, any 2-terminal impedance $Z(\lambda) = \frac{1}{\mu}Z^*(\mu)$, where $\lambda = \mu^2$ and Z^* relates to the corresponding *LC* network. Similarly any admittance $Y(\lambda) = \mu Y^*(\mu)$.‡

Now $Z^*(\mu)$ has simple alternating poles and zeros on the imaginary μ-axis including a pole or zero at $\mu = 0$. Thus $Z(\lambda)$ has corresponding poles and zeros alternating on the negative-real λ-axis, $\lambda \leqslant 0$, the first (at or nearest to $\lambda = 0$) being a pole. $Z(\lambda)$ cannot have a zero at $\lambda = 0$ or a pole at $\lambda = \infty$. Similar results apply to an admittance $Y(\lambda)$ (with "pole" and "zero" interchanged). Moreover, if

$$Z^*(\mu) = \frac{k_0^*}{\mu} + k_\infty^*\mu + \sum_r \frac{2k_r^*\mu}{\mu^2 + \omega_r^2}$$

then $$Z(\lambda) = \frac{k_0}{\lambda} + k_\infty + \sum_r \frac{k_r}{\lambda + \omega_r^2} \quad . \quad . \quad . \quad (33)$$

where $k_0 = k_0^*$, $k_\infty = k_\infty^*$, $k_r = 2k_r^*$. Thus all the residues of $Z(\lambda)$ are positive.

For an admittance $Y(\lambda)$, however, it is Y/λ, not Y, that has the form of eqn. (33). Thus Y has non-negative "coefficients" k_0, etc., but its residues, $-k_r\omega_r^2$ and k_∞, are all negative except the last.

If the same transformation is made to a 4-terminal *RC* network, $z_{ij}(\lambda) = \frac{1}{\mu}z_{ij}^*(\mu)$, $y_{ij}(\lambda) = \mu y_{ij}^*(\mu)$, where z_{ij}^* and y_{ij}^* relate to the corresponding *LC* network. The previous results apply, of course, to the zeros, poles and residues of $z_{11}(\lambda)$, $z_{22}(\lambda)$ and of $y_{11}(\lambda)$, $y_{22}(\lambda)$. Moreover, the poles of $z_{12}(\lambda)$ must be common poles of z_{11} and z_{22}, since those of $z_{12}^*(\mu)$ are common to z_{11}^* and z_{22}^*. Also, the residues of the $z_{ij}(\lambda)$ at any pole are twice the residues of the $z_{ij}^*(\mu)$ at either of the corresponding conjugate imaginary poles (or equal, in the case of poles at $\lambda = 0$). It follows by eqn. (26) that the determinant of the residues of the $z_{ij}(\lambda)$ is non-negative.* Similarly for the $y_{ij}(\lambda)$.

The open-circuit voltage gain from side 1 to side 2 is

$$T = \left(\frac{v_2}{v_1}\right)_{i_2=0} = \frac{z_{12}}{z_{11}}$$

which, by the above, can have poles only at the zeros of $z_{11}(\lambda)$. We may therefore state, using eqns. (28) and (30), that for a 3-terminal *RC* network:

(*a*) The poles of T are distinct negative numbers, and (*b*) $0 \leqslant T \leqslant 1$ for $\lambda \geqslant 0$, with $T = 0$ or 1 possible only for λ equal to zero or infinity, unless the special configuration mentioned in Theorem 1 occurs.

For a 4-terminal *RC* network, (*a*) holds as before, while (*b*) is replaced by (*b'*) $-1 \leqslant T \leqslant 1$ for $\lambda \geqslant 0$, with $T = \pm 1$ possible only for λ equal to zero or infinity, unless the special configuration mentioned in Theorem 2 occurs.

All the conditions in Theorems 1 and 2 of Reference 3 follow immediately from (*a*), (*b*) and (*b'*) (although the special configurations are not mentioned there). The present enunciation is of course much simpler.

Results similar to (*a*), (*b*) and (*b'*) apply to the current gain, τ, in any branch. Furthermore, all the results of the present Section apply equally to *LR* networks without mutual inductance, if impedances are replaced by admittances and conversely.

(9) REFERENCES

(1) CAUER, W.: "Die Verwirklichung von Wechselstrom-widerständen vorgeschriebener Frequenzabhängigkeit," *Archiv für Elektrotechnik*, 1927, **17**, p. 355.

(2) GUILLEMIN, E. A.: "Synthesis of *RC*-networks," *Journal of Mathematics and Physics*, 1949, **28**, p. 22.

(3) FIALKOW, A., and GERST, I.: "The Transfer Function of General Two-Terminal-Pair *RC* Networks," *Quarterly of Applied Mathematics*, 1952, **10**, p. 113.

(4) GUILLEMIN, E. A.: "Communication Networks, II" (Wiley, New York, 1935).

(5) ROUTH, E. J.: "Advanced Rigid Dynamics" (Macmillan, 1892. Fifth edition).

(6) FOSTER, R. M.: "A Reactance Theorem," *Bell System Technical Journal*, 1924, **3**, p. 259.

(7) INGRAM, W. H., and CRAMLET, C. M.: "On the Foundations of Electrical Network Theory," *Journal of Mathematics and Physics*, 1944, **23**, p. 134.

(8) SYNGE, J. L.: "The Fundamental Theorem of Electric Networks," *Quarterly of Applied Mathematics*, 1951, **9** p. 113.

(9) PERCIVAL, W. S.: "The Solution of Passive Electrical Networks by means of Mathematical Trees," *Proceedings I.E.E.*, Paper No. 1492 R, May, 1953 (**100**, Part III p. 143).

(10) KIRCHHOFF, G.: "Über die Auflösung der Gleichungen auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird," *Annalen der Physik und Chemie*, 1847, **72**, p. 497.

(11) TALBOT, A.: "A New Method of Synthesis of Reactance Networks," *Proceedings I.E.E.*, Monograph No. 77 R October, 1953 (**101**, Part IV, p. 73).

† If the numerator and denominator have common factors, this form may not persist after their cancellation.

‡ This technique is described by CAUER, W.: "Theorie der Linearen Wechselschaltungen" (Leipzig, 1941), p. 204; also by BODE, H. W.: "Network Analysis and Feedback Amplifier Design" (Van Nostrand, New York, 1945), p. 214.

* Guillemin[2] "proves" this result by stating that $z_{11} + 2xz_{12} + x^2z_{22}$ is an R impedance function for any real x. He gives no justification for this statement, bu presumably intends it to follow by an argument similar to that on page 216 of h book,[4] where it is used for dealing with *LC* networks. This argument, however, i invalid when applied to *RC* networks, since it makes use of ideal transformers.

(12) BRUNE, O.: "Synthesis of a Finite Two-Terminal Network whose Driving-Point Impedance is a Prescribed Function of Frequency," *Journal of Mathematics and Physics*, 1931, **10**, p. 191.

(13) FEUSSNER, W.: "Über Stromverzweigung in netzförmigen Leitern," *Annalen der Physik*, 1902, **9**, p. 1304; also *ibid.*, 1904, **15**, p. 385.

(14) CAUER, W.: "Über Funktionen mit positivem Realteil," *Mathematische Annalen*, 1932, **106**, p. 369.

(10) APPENDIX

(10.1) Direct Evaluation of Network Parameters

In a paper already mentioned,[10] and strangely neglected for more than a century, Kirchhoff not only laid the foundations of the topology of networks, but gave a rule by which the numerators and common denominator of the branch currents in a resistance network can be written down, term by term, by a topological inspection of the network. An interesting alternative statement of Kirchhoff's results was given by Feussner,[13] who pointed out that Kirchhoff's work seemed then (1902) to have been forgotten for over fifty years, and he showed how to reduce the work involved in writing down the terms. (Percival's work[9] is in effect an extension of Feussner's.) Apart from a slight reference by Ingram[7] to the topological elements in Kirchhoff's work, the author has seen no other mention of the results except by Cauer,[14] who independently proved them. In this Appendix, we state Kirchhoff's results, and show how to deduce expressions for the network parameters α, β, etc.

Kirchhoff's rule may be stated as follows:

If a voltage generator V_i is inserted into a branch of a connected resistance network with n nodes and b branches, and I_h is the current in any branch R_h,

$$\frac{I_h}{V_i} = \frac{N_h}{D} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where $D = \Sigma R_{a(1)}R_{a(2)} \ldots R_{a(M)}$, $N_h = \Sigma \pm R_{b(1)}R_{b(2)} \ldots R_{b(M-1)}$. Here, $M = b - n + 1$, and is the total number of independent meshes; $[R_{a(1)}, \ldots, R_{a(M)}]$ represents any set of M branches such that, after their removal from the network, no closed circuit remains; and $[R_{b(1)}, \ldots, R_{b(M-1)}]$ represents any set of $M - 1$ branches such that after their removal just one closed circuit remains, which includes both V_i and R_h, the associated sign being positive or negative according as the pre-assigned sense of I_h is or is not in agreement with the polarity of V_i in this circuit. In D and N_h summation is over all possible sets.

Now suppose a voltage V_1 is applied externally, through a resistance R_1, to a 3- or 4-terminal resistance network with parameters α, β, γ and δ, and let R_2 be an external load resistance, as shown in Fig. 9.

Then it follows easily from eqns. (16) that

$$\frac{V_1}{I_2} = \alpha R_1 + \beta + \gamma R_1 R_2 + \delta R_2 \quad . \quad . \quad . \quad (34)$$

Now let the inner network have m independent meshes: the complete network has $M = m + 2$. We shall use the symbol Π_t (where $t \geqslant m$) to denote the product of a set of t resistances $R_{a(1)}, \ldots, R_{a(t)}$ in the inner network whose removal destroys all closed circuits in it, and has extra effects as indicated in each case. By applying Kirchhoff's rule to I_2, and separating the terms of D into four groups according to whether they contain both, one or other, or neither of R_1, R_2, it follows by eqn. (34) that

$$(\alpha, \beta, \gamma, \delta) = (a, b, c, d)/g$$

where:

$a = \Sigma\Pi_{m+1}$, where (after removal of the resistances involved in Π_{m+1}) no path 1′1, i.e. between nodes 1′ and 1, remains (in the inner network);

$b = \Sigma\Pi_{m+2}$, where neither a path 1′1 nor a path 2′2 remains;

$c = \Sigma\Pi_m$, where just one path 1′1 remains;

$d = \Sigma\Pi_{m+1}$, where just one path 1′1 remains, if 2′ and 2 are short-circuited;

$g = \Sigma \pm \Pi_{m+1}$, where just one path 12 and one path 1′2′, or else one path 1′2 and one path 12′, remain (and the sign is correspondingly positive or negative).

We can analyse these expressions further: suppose

$p = \Sigma\Pi_{m+1}$, where just one path 12 and one path 1′2′ but no path 1′1 remains;

$q = \Sigma\Pi_{m+1}$, where just one path 1′2 and one path 12′ but no path 1′1 remains;

$r = \Sigma\Pi_{m+1}$, where no path 1′1, no pair of paths 12 and 1′2′, and no pair 1′2 and 12′ remains;

$s = \Sigma\Pi_{m+1}$, where just one path 1′1, but not a pair 12 and 1′2′, nor a pair 1′2 and 12′ remains.

It is clear that these four groups of terms are mutually exclusive, and that:

$$a = p + q + r$$
$$d = p + q + s$$

and

$$g = p - q.$$

Moreover, in the case of a 3-terminal network, where 1′ and 2′ coincide, clearly $q = 0$.

The y_{ij} and z_{ij} can be expressed immediately in terms of these various sums through the relations in eqn. (17).

In conclusion, it should be noted that the three gain theorems, and all the results of Section 4, are simple consequences of the formulae just obtained; and that all the results of this Appendix apply equally to general networks without mutual inductance if the branch resistances R_h are replaced by $Z_h(\lambda)$.

621.317.742 : 621.372.8

Monograph No. 119 R
Jan. 1955

A DIRECT-READING WAVEGUIDE STANDING-WAVE DETECTOR FOR USE AT LOW POWER LEVELS

By H. V. SHURMER, M.Sc., Ph.D., Associate Member.

(*The paper was first received* 24*th June, and in revised form* 14*th October*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in January*, 1955.)

SUMMARY

A method is described of directly measuring voltage standing-wave ratios (v.s.w.r.) at a fixed input power and frequency by means of a directional coupler, in which the errors introduced by the coupler are compensated. The directional coupler is used to measure the power reflected by the load, which is a function of its v.s.w.r.

The main purpose of this development has been to afford a means of directly measuring the v.s.w.r. of low-level detector crystals. In this application it is necessary to have an input power level of not more than about 5μW, below which the v.s.w.r. is substantially constant, and it is required to measure v.s.w.r.'s of the order of 0·80 to an accuracy of $\pm$0·01.

Unless the coupler has infinite directivity and one arm is perfectly matched, an error is introduced into the measurement which depends on the phase of the voltage wave reflected by the load. Since it is difficult to obtain a tightly-coupled directional coupler suitable for use at low power levels (of the order of 5μW) in which the error due to imperfect directivity is negligible, the treatment adopted is to cancel out the voltage waves which produce the error.

An instrument developed on these principles is used to measure the v.s.w.r. of crystal rectifiers at 3·20cm over a range of power levels from 5 to 650μW. The accuracy is better than $\pm$0·01 from 0·90 to zero v.s.w.r.

(1) INTRODUCTION

In measuring the voltage standing-wave ratios (v.s.w.r.) of detector crystals the input power level must not be more than about 5μW, and it is required to measure v.s.w.r.'s of the order of 0·80 to an accuracy of $\pm$0·01.

The paper describes an instrument which has been developed to give direct readings of v.s.w.r. over a narrow bandwidth centred at 3·20cm, at power levels down to 5μW. It will be understood that this wavelength can be varied by a simple tuning arrangement, and, moreover, that the method is not limited to the X-band but is applicable over the whole frequency band associated with waveguides.

The basic equipment required for the measurement of v.s.w.r. by reflected power is shown in Fig. 1(*a*). A klystron oscillator feeds r.f. power modulated at 3kc/s via a wavemeter, variable attenuator and directional coupler to the load. The reflected power from the load is sampled by a crystal detector in one arm of the auxiliary waveguide of the coupler, the other arm of the auxiliary waveguide being terminated in a matched load. The output from the crystal consists of a square wave of the modulation frequency, the fundamental of which is amplified and detected by the selective amplifier and appears on the output meter. The meter can be calibrated directly in terms of the v.s.w.r. of the load provided that some known value of v.s.w.r. is available as a reference, and this is most conveniently provided by substituting a short-circuited termination for the load.

To achieve the required sensitivity from a directional coupler it is necessary that the coupling factor shall be high, of the order of 10dB, so that the gain of the selective amplifier may be sufficiently low for noise troubles to be avoided.

It is extremely difficult to make a directional coupler with a high coupling factor which also has a directivity good enough to achieve the required accuracy—directivity being defined as the ratio of power appearing at the correct outlet to power appearing at the incorrect outlet. This is particularly true if it is required that the coupler shall be used at some predetermined wavelength, on account of the greater difficulty in predetermining the wavelength at which the maximum directivity occurs. A directivity rather better than 40dB would be required to obtain the desired accuracy.[1]

In addition to the error arising from imperfect directivity a further error occurs if there is an imperfectly matched termination at the arm opposite to the detector crystal. This will be serious unless the match is better than 0·99 v.s.w.r.

On account of the error-voltage waves combining in the detector arm with the wave reflected from the load, the magnitude of the detected signal varies with the phase angle of this reflection, and the variation is the more serious the better the v.s.w.r. of the load.

It is shown later how it is possible to cancel out the error-voltage waves by means of a voltage wave derived from a mismatching device arranged to reflect a fraction of the incident wave with the appropriate magnitude and phase. The detected signal then depends solely on the power reflected from the load, and is independent of the phase angle of the reflection.

(2) CANCELLATION OF ERROR-VOLTAGE WAVES

Let us consider an uncorrected directional coupler which is subject to the errors described and is set up for v.s.w.r. measurements as in Fig. 1(*a*).

Let the gain of the selective amplifier be set so that a certain output-meter reading corresponds to some chosen value of v.s.w.r. to which a variable load has been set. This load may conveniently consist of a series combination of phase shifter, single-stub mismatcher and matched termination.

If the phase of the load reflection is varied, it will be found that, to maintain the output-meter reading at its original value, the v.s.w.r. must also be altered. On plotting the locus of the load admittance for a constant output-meter reading with respect to any arbitrary reference plane on a Smith admittance diagram,[2,3] the result will correspond to a circle offset from the centre of the diagram by a certain conductance and susceptance.

To obtain the plot of load admittance for constant output-meter reading it is necessary after each adjustment of the load to remove the entire load combination and measure its admittance using a conventional moving-carriage standing-wave detector.

Fig. 2 shows such a locus obtained experimentally for a particular directional coupler on a partial Smith diagram. It will be clear that, if to each one of the readings a fixed admittance is added, the locus corresponding to constant output-meter reading will be a circle of constant standing-wave ratio, which will correspond to an ideal directional coupler.

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Shurmer is with the British Thomson-Houston Co., Ltd.

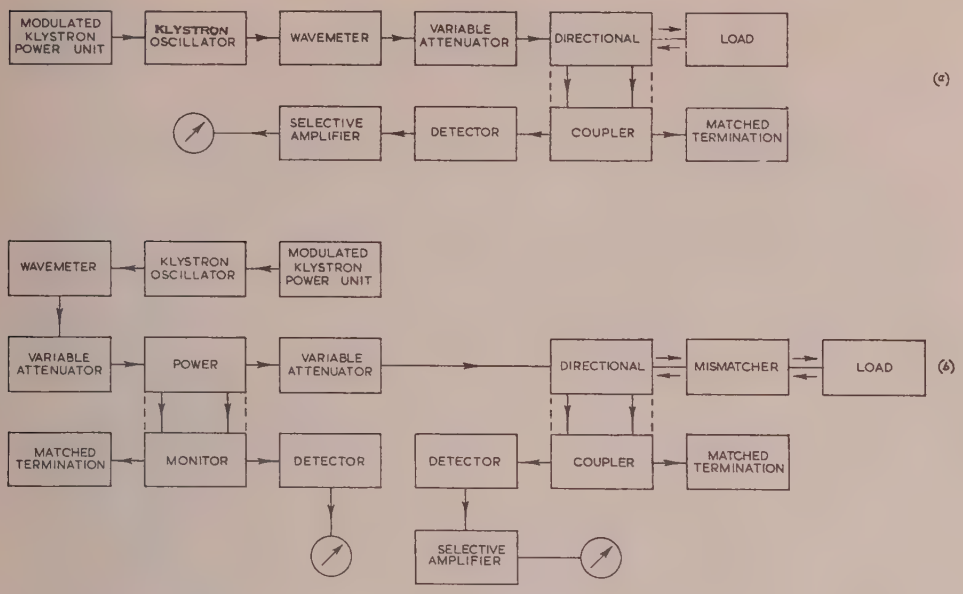


Fig. 1.—Block schematic of direct-reading v.s.w.r. equipment.
(*a*) Basic details. (*b*) Complete equipment.

Furthermore, with this added admittance the locus corresponding to any other meter reading will be a circle of constant v.s.w.r., and conversely, the meter reading will depend only on the v.s.w.r. of the load. Thus the error-voltage waves will be cancelled. It follows that if the load is a perfectly matched termination the output meter will read zero. This leads to a convenient method of achieving the desired compensation.

It will be clear that a fixed admittance to compensate for the errors can be inserted immediately in front of the load in the form of a mismatcher, since both the error-voltage waves and the compensating wave will be proportional to the incident wave. It is also conceivable that the errors could be compensated by a mismatcher inserted before the matched termination in one arm of the coupler, although whether sufficient correction could be achieved by this means would depend on the magnitude of the errors to be compensated for, and on the value of the coupling factor.

A convenient form of mismatcher consists of a four-stub device in which the stub screws are situated one-eighth of a guide wavelength apart. The variation in admittance produced by introducing each of the stubs in turn when such a device is placed before a matched termination is shown in Fig. 3 for an arbitrary reference plane and for standing-wave ratios between 0·5 and unity.

With respect to the plane of each stub the admittance locus follows a semicircle of increasing positive susceptance and unity conductance. Owing to the spatial separation of the stubs the loci are spaced round the diagram at intervals of one-eighth of a wavelength. It will be seen that any value of mismatch can be achieved by using two of the four stubs.

In correcting for errors in the coupler the most convenient procedure is therefore to set up the equipment using as load a good matched termination joined to the coupler via a four-stub mismatcher in which the stubs are initially withdrawn. On account of the inherent errors the amplifier output meter will indicate some finite reading. This reading can be reduced to a minimum by introducing to the correct depth in turn either the first or third stub and the second or fourth stub, the gain of the amplifier being increased to make the minimum position more easily detectable.

When the minimum reading is obtained the two required stubs are locked in position and the other two holes are sealed. This mismatcher is used thereafter as an integral part of the coupler. If the matched termination is removed and the equipment is set up using some arbitrary load, the output-meter reading will be independent of the inherent errors in the coupler.

(3) METHOD OF CALIBRATION

By the definition of voltage standing-wave ratio the reflection factor of the load, r, is given by

$$|r| = \left|\frac{V_r}{V_i}\right| = \left(\frac{1-\rho}{1+\rho}\right) \quad . \quad . \quad . \quad . \quad (1)$$

where V_i = Incident voltage wave at load.
V_r = Reflected voltage wave at load.
ρ = V.S.W.R. of load.

Hence the reflected power, P_r, may be written in terms of the incident power, P_i, as

$$P_r = P_i\left(\frac{1-\rho}{1+\rho}\right)^2 \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Thus for constant input power the output power is a simple function of the v.s.w.r. If the device for detecting the reflected power is a square-law crystal feeding into an amplifier with a linear output meter, the meter reading will be proportional to the reflected power. If R is the meter reading and C is a constant of proportionality

$$R = C\left(\frac{1-\rho}{1+\rho}\right)^2 \quad . \quad . \quad . \quad . \quad . \quad (3)$$

If the gain of the amplifier is set so that a certain meter reading corresponds to some chosen v.s.w.r., C may be evaluated for this setting, and hence the meter reading corresponding to any other v.s.w.r.

In an instrument which has been developed on the above principles primarily for measurements on crystal rectifiers, a

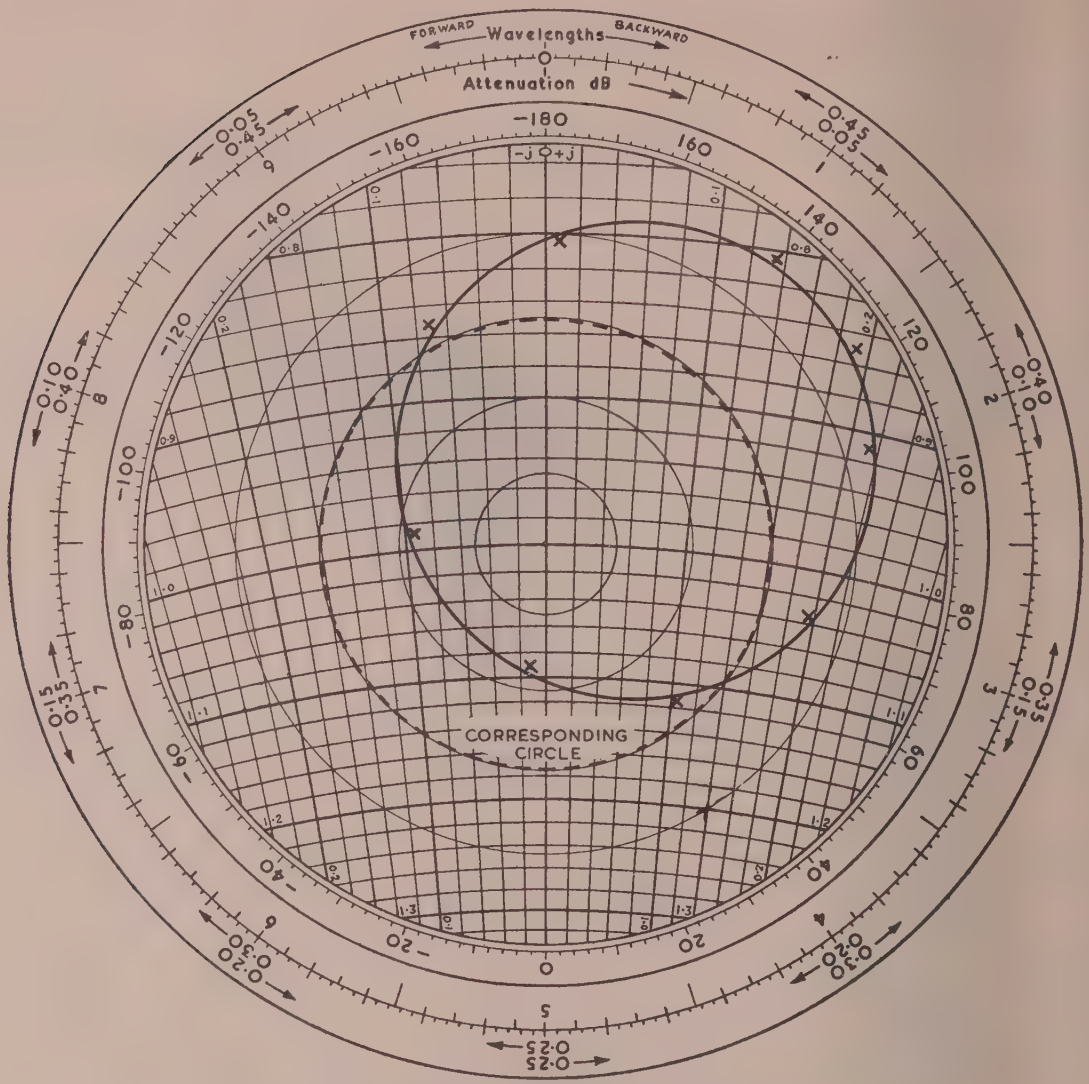


Fig. 2.—Admittance locus for constant output-meter reading.
Partial Smith diagram. V.S.W.R.—0·75: 1.
(Full diagram: 31·5 in diameter.)

convenient value at which to set the gain is such that a v.s.w.r. of 0·80 corresponds to 0·4 times the full-scale meter reading. The meter scale can then conveniently be used to read off values of v.s.w.r. between 0·90 and 0·70, the latter corresponding to full scale on the meter.

In order to increase the range it is convenient to insert a fixed attenuator step in series with the selective amplifier. A 10dB step will extend the range to 0·52 v.s.w.r., 20dB to 0·29 v.s.w.r., and a short-circuit would give a reading of 0·324 times full scale with 40dB attenuation inserted. The most convenient method of setting up the equipment is to adjust the gain of the amplifier until the above reading is obtained with a short-circuited termination. The stepped attenuator may conveniently consist of an inter-stage tapped resistance chain within the selective amplifier.

(4) RANGES OF OPERATION

With the equipment set up as described, the various ranges which result are given in Table 1.

Table 1

RANGES OF V.S.W.R. CORRESPONDING TO DIFFERENT ATTENUATIONS

Range	Attenuation	V.S.W.R.	
		Maximum	Minimum
	dB		
1	0	0·90	0·70
2	10	0·80	0·52
3	20	0·60	0·29
4	30	0·30	0·01
5	40	zero at 0·324 times full scale	

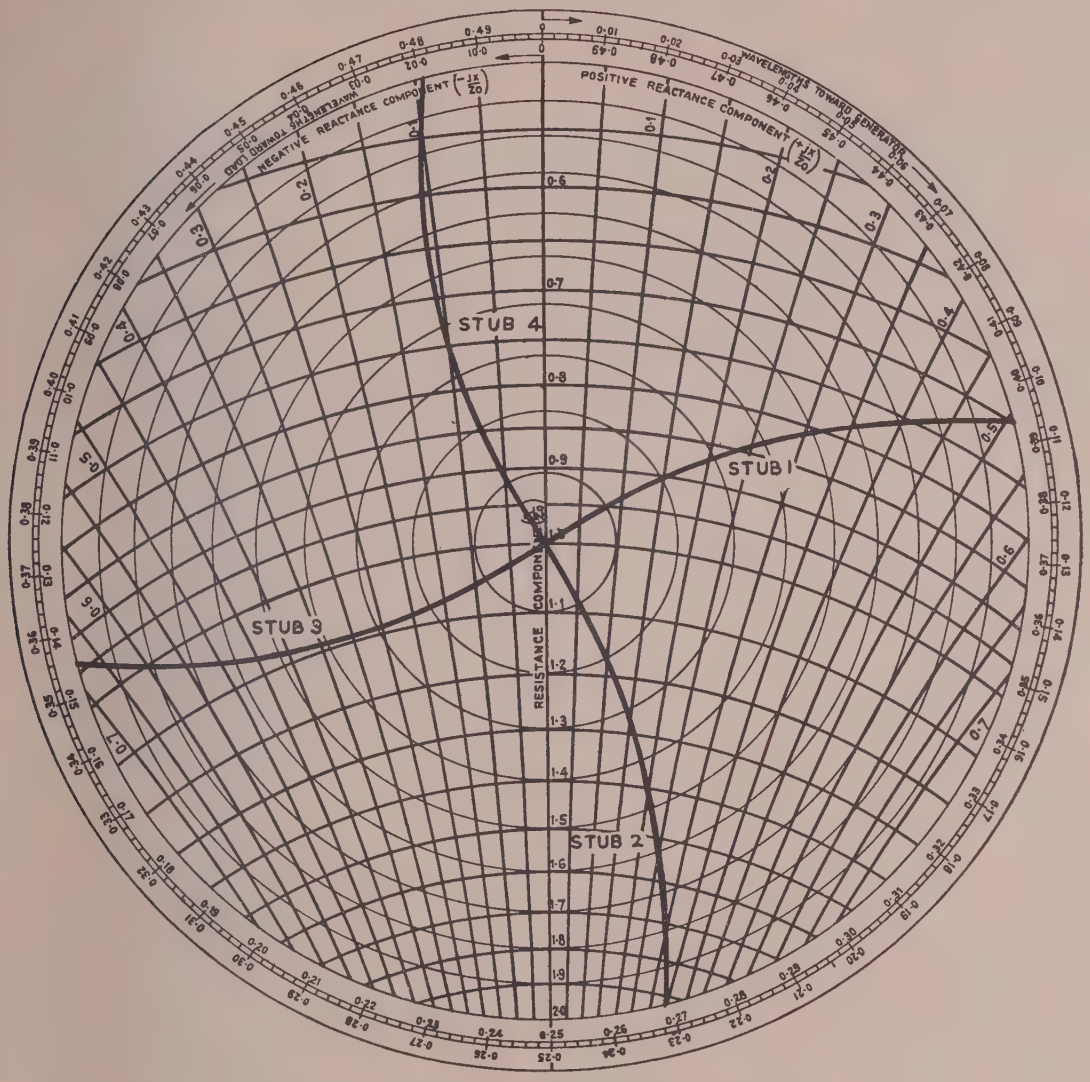


Fig. 3.—Admittance loci corresponding to mismatcher stubs.
Partial Smith diagram. V.S.W.R.—0·50: 1.
(Full diagram: 13·5 in diameter.)

The equipment can be used down to at least an input power level of 5μW. The maximum power at the detector arm occurs with a load of zero v.s.w.r. For a 10dB coupler this will be 0·5μW with an input power of 5μW at the load, and this is well within the square-law region of any detector crystal.

At much higher levels care should be taken that the detector crystal is not operated outside the region over which it has a square-law characteristic, which will depend on the particular crystal, but an upper limit of 5μW may be taken as a guide. In order to ensure this it may be necessary to use a waveguide attenuator in front of the detector crystal. The detector should be reasonably well matched to the holder, a condition which can be achieved with the aid of a four-stub mismatcher such as has been described.

(5) ACCURACY

A comparison of experimental results has been made on a range of 18 shielded-type crystal rectifiers having a variation in v.s.w.r. between 0·89 and 0·29. Two readings obtained on each crystal, using the equipment which has been described, were compared with two similar measurements given by a moving-carriage standing-wave detector. Between readings the crystals were removed and re-inserted in the holder.

It was found that the r.m.s. variation between the pairs of readings on the equipment described was 0·0028 compared with an r.m.s. variation of 0·0024 using the moving-carriage standing-wave detector. The r.m.s. variations between the first set of measurements on the direct-reading equipment and the two sets of measurements on the moving-carriage standing-wave detector were 0·0048 and 0·0052, respectively. It is thus seen that the discrepancy between readings obtained by the two methods is of the same order as the accuracy to be expected of either method.

The above results were obtained at 3·20cm at a power level of 650μW and using a 20dB waveguide attenuator in front of the detector crystal in the direct-reading method. On removing the waveguide attenuator the equipment could be used at 5μW input power level with a similar detected signal, for any value of v.s.w.r. as at 650μW and with the gain setting of the selective

amplifier at nearly the same value. Thus similar conditions, so far as the accuracy is concerned, were obtainable at 5 μW.

(6) OTHER NOTES

The equipment is reasonably stable after an initial warming-up period of half an hour.

The method described has been developed for the production testing of crystal rectifiers, but is also suitable for determining the v.s.w.r. of other microwave equipments.

A useful refinement to the equipment is shown in the complete block schematic of Fig. 1(*b*). The input power is supplied via two variable attenuators separated by a directional coupler which monitors the input power. By means of the first of these attenuators, the detected signal may be set to any convenient value and the second attenuator set to give the required power at the load. Should any variation in source power occur once the equipment has been set up, this may be compensated by varying the first attenuator to return the monitored signal to its original value.

(7) ACKNOWLEDGMENTS

The author wishes to thank and acknowledge the help given to him by Miss J. M. Fennell in the experimental work; he also wishes to thank Mr. L. J. Davies, Director of Research, British Thomson-Houston Co., Ltd., for permission to publish the paper.

(8) REFERENCES

(1) RATCLIFFE, P. M.: "The Directional Coupler as an Impedance Measuring Device," *Marconi Instrumentation*, 1953, **4**, p. 9.

(2) SMITH, P. H.: "Transmission Line Calculator," *Electronics*, 1939, **12**, p. 29.

(3) SMITH, P. H.: "An Improved Transmission Line Calculator," *ibid.*, 1944, **17**, p. 130.

621.376.23 : 621.38

Monograph No. 120 R
Feb. 1955

SIGNAL/NOISE PERFORMANCE OF MULTIPLIER (OR CORRELATION) AND ADDITION (OR INTEGRATING) TYPES OF DETECTOR

By D. G. TUCKER, D.Sc., Ph.D., Member.

(*The paper was first received* 19*th June, and in revised form* 23*rd September*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in February*, 1955.)

SUMMARY

An analysis is made, and numerical results are tabulated, of the processes of addition and multiplication of signals (comprising tone or noise) in the presence of random-noise backgrounds. These processes may be carried out on the input signals, in which case coincidence in time and phase of the signals is required, or on the signals after they have been rectified and smoothed (i.e. on the envelopes), in which case coincidence is required only in time. If the backgrounds are uncorrelated with one another, then the addition or multiplication of two or more scans together results in an increase in detectability of the signal. It is shown that, in general, multiplication is not more advantageous than addition from the point of view of signal/noise ratios, and as it is more difficult to perform in practice, it cannot be universally recommended. But it is undoubtedly better than addition in certain cases, notably those in which long-term stability of adjustment is required and constant or predictable phase relationships permit multiplication of the inputs before rectification.

Some subjective evidence is brought forward and discussed in relation to the difficult question of whether the better signal/noise performance given by a quadratic response as compared with a linear response is really indicative of a higher probability of detection. The conclusions are not clear, but it is possible that this is the case.

LIST OF THE MORE IMPORTANT SYMBOLS

R_1 = Input signal/noise ratio.
V_{n1} = Input noise voltage (r.m.s.).
$V_1 \cos \omega_p t$ = Coherent input signal.
R_{A1}, R_{A2}, R_{B1}, R_{B2} = Detectability criteria defined in Section 2.
V_0 = Output voltage.
V_{n0} = Output noise voltage (r.m.s.).
R_0 = Output signal/noise ratio.
$x \cos (\omega_p - \omega_q)t$, $y \cos (\omega_p - \omega_s)t$ = Typical components of background noise.
$z \cos (\omega_p - \omega)t$ = Typical component of noise signal.

(1) INTRODUCTION

The signal/noise performance of a detector may often be improved by multiplying or adding together two or more "scans" in which the signals coincide in time or are correlated, and in which the noise backgrounds are uncorrelated. The adding process is quite familiar and occurs in radar systems by virtue of the afterglow of a cathode-ray-tube screen; it can also be arranged by means of storage tubes or magnetic-drum or -tape systems, and there is no theoretical limit to the number of additions. In practice, the signal generally moves in the time scale, and so there is a practical limit to the number of scans which can be added with effective coincidence of signal position. The adding process is often referred to as integration—or pulse-to-pulse integration, or scan-to-scan integration—by analogy with the similar results of filtration, which is a true integration process.

Multiplying scans together is a less familiar process, although it occurs in the interferometer techniques used in radio-astronomy[1,2] and elsewhere. In such techniques, the "scans" which are multiplied are those obtained simultaneously on two separate arrays. But obviously successive scans of a single array may be multiplied together if a storage system is used. The multiplying-type detector is often referred to as a correlation detector, because its d.c. output represents the correlation coefficient (or function) of the two input signals.

Multiplying or adding may be performed on single-array systems without storage devices if two or more independent scans can be simultaneously obtained by using, for example, different frequency-bands. Thus it can be seen that the scope for multiplier or addition detectors is quite wide.

Multiplying or adding may generally be done on the incoming signal or after rectification, although when different frequency-bands are involved, as in the previous paragraph, there is no choice but to process the signals after rectification. If the processing is done on the incoming signal, attention must be paid to phasing of the signal.

The purpose of this paper is to examine the signal/noise performance of these processes. No other aspect will be considered in detail. The analysis is based on a simplified representation of noise by a series of small but equal-amplitude tones of different frequencies, and lack of correlation between two noise waveforms is shown by taking different series, each with a quite independent set of frequencies. This representation, while possibly not rigorous, is thought to give no error in the present application. The signal/noise criteria are those defined in a previous paper,[3] and take no particular account of special probability distributions of the noise. As the author is concerned mainly with narrow-band pulse systems—i.e. those with optimum input filtration—in which there is little scope for post-detector integration (i.e. low-pass filtration), there is no discussion of the effect of such integration. However, it is throughout assumed that the h.f. components (e.g. the input frequency-band upwards) are removed from the detector outputs, immediately after rectifying.

The general conclusions reached are that multiplying is not, in general, more effective than adding, and being more difficult has often, therefore, little to recommend it. It does have a slight advantage due to its square-law dynamic response, but this can more simply be obtained by using square-law rectifiers. When done on the unrectified signal, multiplying has the big advantage of giving a final output which contains no d.c. component in the absence of correlated signal. This is very important when the background is liable to vary, since then, in a system using addition followed by rectification, the fluctuating d.c. background may be difficult to bias out, and will often more than offset the advantage in signal/noise performance this system has at low input-signal/noise ratios; but the multiplier, having

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Tucker is in the Royal Naval Scientific Service.

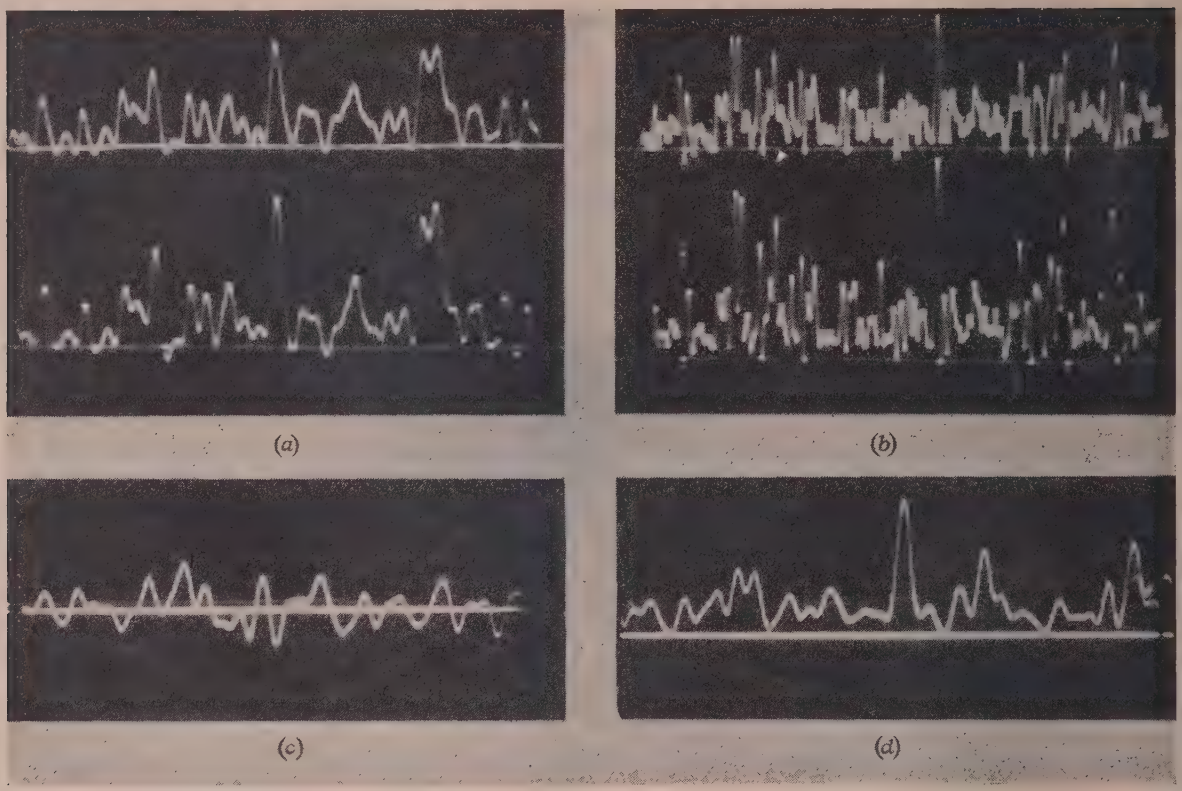


Fig. 1.—Oscillograms of noise waveforms from various detectors.

Outputs are smoothed to leave only envelope frequencies.
(*a*) Outputs of linear (top) and square-law (bottom) detectors for the same input waveform.
(*b*) As (*a*), but with slower time-scan.
(*c*) Output of multiplier (multiplying unrectified signals) when the two noise inputs are uncorrelated.
(*d*) As (*c*), but when noise inputs are completely correlated (i.e. same waveform applied to both inputs).

no d.c. background, will be free from such difficulties. This is the main advantage of the multiplier over the addition system in radio-astronomy. Figs. 1(*a*)–1(*c*) are oscillograms of typical background (white noise) waveforms at the output of a linear detector, square-law detector, and multiplier (multiplying two uncorrelated noise inputs without prior rectification). The zero voltage axis is shown in all three waveforms, and the absence of a d.c. output in the multiplier is clear. For completeness, Fig. 1(*d*) shows the multiplier output when the noise inputs are correlated exactly; a d.c. output now exists.

The analysis is in terms of one addition or one multiplication, but evidently the processes can be repeated as long as the signal pulses can be added in phase or in coincidence, and the end results are easily deduced from the analysis and Tables presented here.

(2) RESULTS

Details of the analysis of the signal/noise performance of the various detector arrangements (except for those previously published[3,4]) are given in the Appendices, and numerical results over the range of R_1 (input-signal/noise ratio) from 0·1 up to 4 are given in Tables 1 and 2.

It should be noted that some of the arrangements analysed give dynamic characteristics which are, at least nominally,* linear; others are basically quadratic, while one or two have a quartic response. Table 1 includes those cases which are nominally linear, and Table 2 those which are quadratic. The cases of quartic-type response (e.g. multiplication of the outputs of square-law rectifiers) are not worked out numerically, since they are not thought to be of very great interest; but their analysis is included in the Appendices, so that all basic information is available should numerical results ever be required.

The criteria of signal/noise performance used are those defined in a previous paper,[3] namely

$$R_{A1} = \frac{\text{(d.c. output due to signal and noise) + (r.m.s. output of l.f. noise when signal present)}}{\text{d.c. + r.m.s. value of l.f. output when signal absent.}}$$

$$R_{A2} = \frac{\text{Root mean square of d.c. plus l.f. components when signal present}}{\text{Root mean square of d.c. plus l.f. components when signal absent}}$$

$$R_{B1} = \frac{\text{Change in direct current on application of signal}}{\text{l.f. noise when signal present}}$$

$$R_{B2} = \frac{\text{Change in direct current on application of signal}}{\text{l.f. noise when signal absent}}$$

The coherent detector[4] has been included among the linear detectors (it is obviously applicable only to coherent-tone signals and not to noise signals) for comparison. The outputs of two coherent detectors with uncorrelated backgrounds can evidently

* "Nominally linear" because at low signal/noise ratios the behaviour of the linear rectifier is effectively quadratic.

Table 1

Detectors with Linear Dynamic Response

R_1	R_{A1}							R_{A2}						
	A	B	C	D	A	B	C	A	B	C	D	A	B	C
	X	X	X	X	Y	Y	Y	X	X	X	X	Y	Y	Y
4	3·53	4·74	3·75	6·66	4·12	5·75	4·87	4·12	5·75	4·39	5·75	4·12	5·75	4·55
2	2·08	2·68	2·15	3·83	2·24	3·0	2·68	2·24	3·0	2·34	3·0	2·24	3·0	2·49
1	1·39	1·62	1·42	2·41	1·41	1·73	1·63	1·41	1·73	1·45	1·73	1·41	1·73	1·54
0·5	1·12	1·20	1·12	1·70	1·12	1·22	1·20	1·12	1·22	1·13	1·22	1·12	1·22	1·16
0·2	1·02	1·04	1·02	1·28	1·02	1·04	1·04	1·02	1·04	1·02	1·04	1·02	1·04	1·027
0·1	1·005	1·01	1·005	1·14	1·005	1·01	1·01	1·005	1·01	1·005	1·01	1·005	1·01	1·007

R_1	R_{B1}							R_{B2}						
	A	B	C	D	A	B	C	A	B	C	D	A	B	C
	X	X	X	X	Y	Y	Y	X	X	X	X	Y	Y	Y
4	4·5	6·8	6·42	5·65	1·52	1·65	1·22	6·9	10·4	9·69	5·65	6·24	9·5	8·44
2	1·85	2·9	2·59	2·82	1·11	1·33	0·86	2·7	4·4	3·8	2·82	2·48	4·0	3·34
1	0·65	1·1	0·90	1·41	0·59	0·85	0·50	0·85	1·55	1·21	1·41	0·83	1·46	1·12
0·5	0·22	0·4	0·29	0·71	0·214	0·36	0·23	0·22	0·44	0·32	0·71	0·24	0·44	0·32
0·2	0·039	0·078	0·053	0·282	0·04	0·08	0·05	0·039	0·078	0·054	0·282	0·04	0·08	0·054
0·1	0·010	0·02	0·013	0·141	0·01	0·02	0·013	0·010	0·02	0·014	0·141	0·01	0·02	0·013

A = Linear rectifier.
B = Addition followed by linear rectifier.
C = Addition after linear rectifiers.
D = Coherent detector.

X for coherent signal.
Y for noise signal.

Table 2

Detectors with Quadratic Dynamic Response

R_1	R_{A1}								R_{A2}							
	E	F	G	H	E	F	G	H	E	F	G	H	E	F	G	H
	X	X	X	X	Y	Y	Y	Y	X	X	X	X	Y	Y	Y	Y
4	11·37	28·4	14·6	12·34	17·0	45·8	22·1	21·69	12·7	28·3	16·84	14·28	17·0	32·5	22·0	21·47
2	4·00	8·63	4·72	4·17	5·0	12·0	6·09	6·52	4·12	6·4	5·00	4·44	5·0	8·55	6·07	6·45
1	1·87	3·13	1·98	1·90	2·0	3·64	2·17	2·49	1·87	2·22	2·00	1·91	2·0	2·63	2·17	2·46
0·5	1·24	1·58	1·25	1·24	1·25	1·62	1·27	1·39	1·24	1·275	1·25	1·24	1·25	1·32	1·27	1·37
0·2	1·04	1·095	1·04	1·04	1·04	1·10	1·04	1·06	1·04	1·04	1·04	1·04	1·04	1·045	1·038	1·06
0·1	1·01	1·025	1·01	1·01	1·01	1·025	1·01	1·01	1·01	1·01	1·01	1·01	1·01	1·01	1·010	1·015

R_1	R_{B1}								R_{B2}							
	E	F	G	H	E	F	G	H	E	F	G	H	E	F	G	H
	X	X	X	X	Y	Y	Y	Y	X	X	X	X	Y	Y	Y	Y
4	2·8	3·94	3·88	3·94	0·94	0·97	1·09	0·80	16·0	22·6	25·33	22·6	16·0	22·6	25·5	22·6
2	1·33	1·89	1·79	1·89	0·80	0·88	0·97	0·64	4·0	5·66	6·05	5·66	4·0	5·66	6·18	5·66
1	0·60	0·80	0·76	0·80	0·50	0·635	0·65	0·44	1·0	1·42	1·39	1·42	1·0	1·41	1·44	1·414
0·5	0·25	0·29	0·26	0·29	0·20	0·278	0·26	0·22	0·25	0·35	0·319	0·35	0·25	0·354	0·33	0·354
0·2	0·04	0·054	0·049	0·054	0·038	0·054	0·05	0·052	0·04	0·057	0·051	0·057	0·04	0·057	0·052	0·057
0·1	0·01	0·014	0·013	0·014	0·010	0·014	0·013	0·014	0·01	0·014	0·013	0·014	0·01	0·014	0·013	0·014

E = Square-law rectifier.
F = Multiplier.
G = Multiplier after linear rectifiers.
H = Addition after square-law rectifiers.

X for coherent signal.
Y for noise signal.

be added together or multiplied just as with other detectors, and in the case of the R_B criteria, an addition gives an improvement of $\sqrt{2}$ (i.e. 3 dB) just as with linear or square-law rectifiers. But these cases are not included in the Tables.

In Table 1 it can be observed that whichever criterion is used, adding before rectifying is generally better than adding after rectifying. On criteria R_{B1} and R_{B2} the improvement relative to a single rectifier is a factor of about $\sqrt{2}$ at high signal/noise ratios; but at low signal/noise ratios, adding before rectifying profits from the square-law nature of the linear-detector response at low ratios and doubles the values of R_{B1} and R_{B2}, while adding after rectification merely continues to give the $\sqrt{2}$ improvement.

In Table 2 it can be observed that:

(*a*) On criteria R_{A1} and R_{A2}, multiplying the input signals together is far better than multiplying after rectification—in fact, the latter is little better than a single output.

(*b*) On criteria R_{B1} and R_{B2}, there is little to choose between multiplying the input signals together, multiplying after rectification, or adding after square-law rectifiers—except R_{B1} with a noise-signal, which is a case of no practical significance, as R_{B1} is not a suitable criterion for noise signals—but all three cases are about $\sqrt{2}$ times better than a single square-law rectifier. Clearly, adding before square-law rectification gives a doubling of the values of R_{B1} and R_{B2} (this case is not tabulated, being obvious) and is therefore superior to all other cases.

The significance of a comparison between Tables 1 and 2 is not immediately clear, and one aspect of it is discussed in the next Section.

(3) LINEAR VERSUS QUADRATIC DYNAMIC RESPONSE

It is clear from Tables 1 and 2 that those detection circuits with a quadratic dynamic response have considerably higher values of R_{A1}, R_{A2} and R_{B2} at the upper ranges of R_1, and noticeably higher values at the lower ranges, than those circuits with a linear response. It is tempting to think that this means that better detection is obtained by using circuits with quadratic (or higher-order) response instead of linear circuits. In a previous paper[3] doubt was cast on the truth of this conclusion, and the relationship between subjective probability of detection of a signal against noise and the objective signal/noise ratio and its derived criteria (such as R_A and R_B) has never been clear. There can, of course, be no unique relationship between probability of detection and signal/noise criteria, because subjective detection is a function of many factors, such as display efficiency, operator fatigue, environment, any pattern or shape in the signal as seen on the display, assistance by aural presentation, etc. The real question is whether, for any given set of display and operating conditions, signal/noise criteria can be used to compare the merits of different kinds of detector circuit. They are useful criteria only if they can be so used.

Recently, subjective measurements of threshold-detection performance have been made, and a limited measure of support for the use of signal/noise criteria has been obtained. So far as intensity-modulated displays are concerned, it has been shown —theoretically[8] as well as practically—that contrast, obtained by the use of bias, leads to better detection when the dynamic range of the display, measured in terms of "just-noticeable differences," is limited; criterion R_A (i.e. either R_{A1} or R_{A2}), which has been considered appropriate for intensity-modulated displays, does in fact give higher values when contrast is applied, and to this extent may be considered a reliable criterion. But the problem of the performance of intensity-modulated displays is very complicated and no comprehensive data are yet available.

The problem of detection by A-scans is much easier, and comparative results of linear and square-law detectors using an A-scan display are available. Experimental results obtained for probability of detection of a pulse of 20 millisec duration against "white" noise restricted to a 200 c/s bandwidth, using A-scan cathode-ray-tube display with a scan duration of 250 millisec, and averaged over several hundred observations by several different observers, show no large difference in detectability between single linear and square-law detectors. Fig. 2 shows a

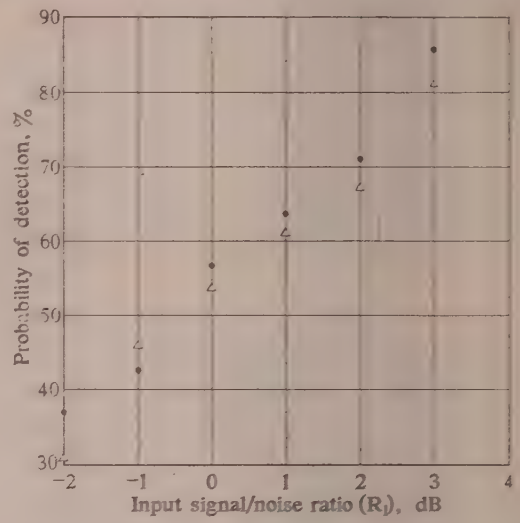


Fig. 2.—Subjective detection results for three operators.
● Square-law detector.
△ Linear detector.

graph of percentage probability of detection against input signal/noise ratio (R_1) for the average of a hundred observations at each value of R_1 by each of three observers on each type of detector, using for each observation a single scan on a long-persistence cathode-ray tube. The pulse, which was of coherent tone, was applied in one of ten positions on the scan, and the observer had to state in which position the pulse appeared. No "nil" record was allowed, and the percentage probability of detection is the actual percentage of correct detections obtained. Results are shown for linear and square-law detectors. It will be seen that over the range of R_1 from −2 to +3 dB there is only a slight difference in favour of the square-law detector. Fig. 3 shows a similar pair of curves based on one hundred observations for each value of R_1 on each type of detector, made by one particular observer (among the three) whose results were more consistent than those of the two other observers; here the difference in favour of the square-law detector is greater. Assuming these results are, in fact, more reliable than the others, and that the straight-line approximations are valid, we shall now try to relate these probabilities of detection to the criterion R_{B2} (which is considered the appropriate one for *A*-scans with short pulses) using the calculations summarized in Tables 1 and 2.

Fig. 4 shows a graph of R_{B2} against R_1 for both linear and square-law detectors. From this and Fig. 3, points can be plotted, as shown in Fig. 5, relating probability of detection to R_{B2}. It will be seen that over the range of probabilities from 30 to 85%, the points lie remarkably closely on one and the same curve, as shown. If the straight-line approximations of Fig. 3 are not used, but instead the actual observed values are taken in constructing Fig. 5, then the points lie equally well on a single curve except for those corresponding to $R_1 = +3$ dB. These results support the hypothesis that for this particular kind of display the relationship between probability of detection and R_{B2} is independent of the type of detector circuit used, so

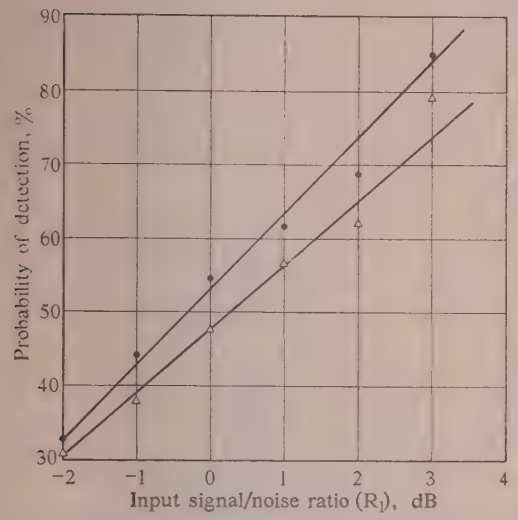


Fig. 3.—Subjective detection results for one more-consistent operator.
● Square-law detector.
△ Linear detector.

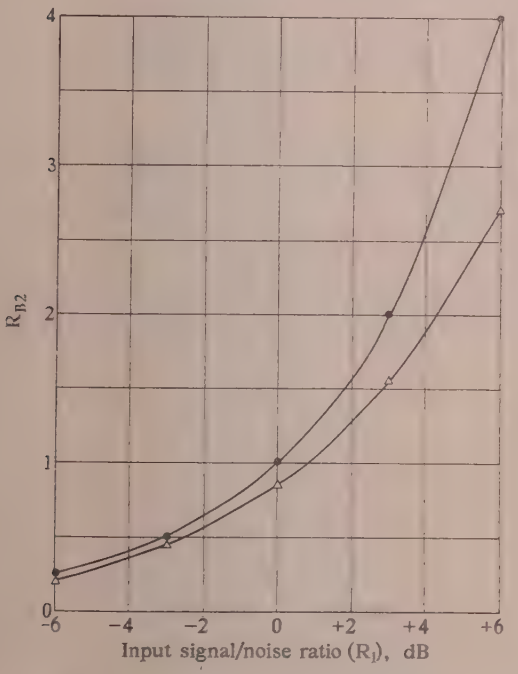


Fig. 4.—Graphs of detection criterion R_{B2} against input signal/noise ratio, R_I.
● Square-law detector.
△ Linear detector.

that the signal/noise criterion may well be perfectly reliable as a means of comparing the performance of different detector circuits. There is, of course, no proof as yet, and the above comparisons have been made only on the simplest cases. Moreover, the data used are not very comprehensive, and are selected as being the most reliable among a larger batch of data, which, taken as a whole, give rather less impressive results. The subjective investigation of detectability is proceeding.

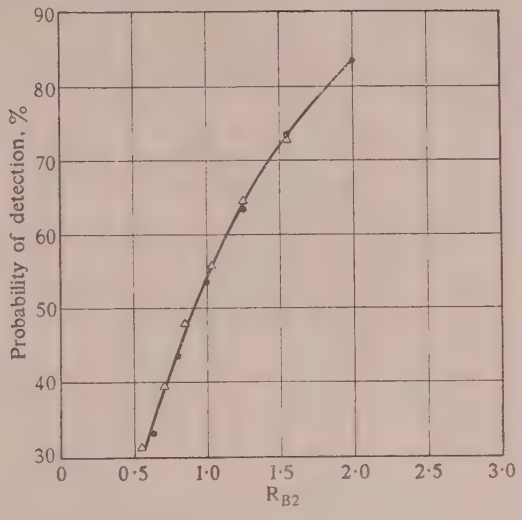


Fig. 5.—Graphs of probability of detection against R_{B2} for the one more-consistent operator.
● Square-law detector.
△ Linear detector.

The matter most open to doubt is whether the divergence of values of R_A and R_B as between linear and square-law responses at large R_I represents any practical difference in detectability. At these larger values of R_I the probability of detection is very high, and consequently it cannot be altered very much by changing from linear to square-law detectors. But it is possible, of course, and even probable, that the greater contrast of the display obtained with the square-law response gives the operator more confidence in his detections, and—partly because of this—reduces operator fatigue. This would lead eventually to better detection under operational conditions. There is, however, no experimental evidence available on this aspect of the problem.

(4) CONCLUSIONS AND ACKNOWLEDGMENTS

The general conclusions are given in the Summary and in the Introduction, and need not therefore be repeated here. More detailed conclusions are given in Section 2, and all numerical results are tabulated in Tables 1 and 2.

The author is glad to acknowledge the very considerable help given by Mr. G. L. Connon in checking the analytical work and in much of the numerical computation, and also by Messrs. J. W. R. Griffiths and B. G. Hunt, who obtained the experimental data on which Section 3 is based. The paper is published by permission of the Admiralty. Any opinions expressed are, of course, personal ones of the author.

(5) REFERENCES

(1) Ryle, M.: "A New Radio Interferometer and its Application to the Observation of Weak Radio Stars," *Proceedings of the Royal Society*, A, 1952, **211**, p. 351.

(2) Hanbury Brown, R., Jennison, R. C., and Das Gupta, M. K.: "Apparent Angular Sizes of Discrete Radio Sources," *Nature*, 1952, **170**, p. 1061.

(3) Tucker, D. G., and Griffiths, J. W. R.: "Detection of Pulse Signals in Noise," *Wireless Engineer*, 1953, **30**, p. 264.

(4) Tucker, D. G.: "The Synchrodyne and Coherent Detectors," *ibid.*, 1952, **29**, p. 184.

(5) Bennett, W. R.: "Response of a Linear Rectifier to Signal and Noise," *Journal of the Acoustical Society of America*, 1944, **15**, p. 164.

(6) JAHNKE, E., and EMDE, F.: "Tables of Functions" (Leipzig, 1938, or New York, 1943).
(7) TUCKER, D. G.: "Linear-Rectifiers and Limiters: Effect of Complex Signals and Noise," *Wireless Engineer*, 1952, **29**, p. 128.
(8) TUCKER, D. G., and GRIFFITHS, J. W. R.: "Detection of Pulse Signals in Noise by Intensity-Modulated Visual Display: Theory in Terms of Just-Noticeable Differences." *ibid.* (in course of publication).

(6) APPENDICES

(6.1) Multiplier-Type Detector with Coherent Signal and Uncorrelated Background Noise

Let the signal on each input be $V_1 \cos \omega_p t$.

Let the noise be made up of a multiplicity of tones of the type $xV_1 \cos (\omega_p - \omega_q)t$ on Input A, and $yV_1 \cos (\omega_p - \omega_s)t$ on Input B.

Thus

$$V_A = V_1[\cos \omega_p t + x_1 \cos (\omega_p - \omega_{q1})t + \ldots + x_n \cos (\omega_p - \omega_{qn})t] \quad . \quad . \quad . \quad . \quad (1)$$

and

$$V_B = V_1[\cos \omega_p t + y_1 \cos (\omega_p - \omega_{s1})t + \ldots + y_n \cos (\omega_p - \omega_{sn})t] \quad . \quad . \quad . \quad . \quad (2)$$

and on multiplying these together and eliminating (by filtration) all high-frequency bands, we are left with the d.c. and modulation-frequency (ω_q) terms thus:

$$\begin{aligned} V_A \times V_B = {} & \tfrac{1}{2}V_1^2[1 + x_1 \cos \omega_{q1} t + \ldots + x_n \cos \omega_{qn} t \\ & + y_1 \cos \omega_{s1} t + \ldots + y_n \cos \omega_{sn} t \\ & + x_1 y_1 \cos (\omega_{q1} - \omega_{s1})t + \ldots + x_n y_1 \cos (\omega_{qn} - \omega_{s1})t \\ & + x_1 y_2 \cos (\omega_{q1} - \omega_{s2})t + \ldots + x_n y_2 \cos (\omega_{qn} - \omega_{s2})t \\ & + \; . \\ & \quad . \\ & \quad . \\ & + x_1 y_n \cos (\omega_{q1} - \omega_{sn})t + \ldots + x_n y_n \cos (\omega_{qn} - \omega_{sn})t] \end{aligned}$$

(n^2 terms)

. . . . . (3)

If we now make all the x's and y's equal—which is reasonable, since we assume both inputs have the same signal/noise ratio—we obtain

$$\text{Signal output} = \tfrac{1}{2}V_1^2 \quad . \quad . \quad . \quad . \quad . \quad (4)$$

$$V_{n0} = \text{r.m.s. noise output} = \frac{1}{\sqrt{2}} \frac{V_1^2}{2}(2nx^2 + n^2x^4)^{\frac{1}{2}} = \tfrac{1}{2}V_1^2 \sqrt{(nx^2)}\left(1 + \frac{nx^2}{2}\right)^{\frac{1}{2}} \quad . \quad (5)$$

Let the input signal/noise ratio be R_1 (on each input), and the output signal/noise ratio be R_0. Then, if V_{n1} = r.m.s. input noise voltage (on each input),

$$R_1 = \frac{V_1}{\sqrt{(2)}V_{n1}} = \frac{1}{\sqrt{(nx^2)}}$$

so that

$$R_0 = R_1 \Big/ \sqrt{\left(1 + \frac{1}{2R_1^2}\right)} \quad . \quad . \quad . \quad . \quad (6)$$

It is useful also to express V_{n0} in terms of R_1, thus:

$$V_{n0} = \frac{1}{2}\frac{V_1^2}{R_1} \sqrt{\left(1 + \frac{1}{2R_1^2}\right)} \quad . \quad . \quad . \quad . \quad (7)$$

We also require to know the output noise voltage when the coherent signal is absent. This cannot be obtained from eqn. (7), which is indeterminate when $V_1 = 0$ so that $R_1 = 0$. But it is clear that in these circumstances, only the "xy" terms in eqn. (3) have existence, and so we obtain

$$V_{n0} \text{ (in absence of signal)} = V_{n1}^2/\sqrt{2} \quad . \quad . \quad . \quad (8)$$

From the above results, we obtain for the pulse signal/noise criteria

$$R_{A1} = \sqrt{2}\left[R_1^2 + R_1\sqrt{\left(1 + \frac{1}{2R_1^2}\right)}\right] \quad . \quad . \quad (9)$$

$$R_{A2} = \sqrt{2R_1^2}\sqrt{\left[1 + \frac{1}{R_1^2}\left(1 + \frac{1}{2R_1^2}\right)\right]} \quad . \quad . \quad (10)$$

$$R_{B1} = R_0 = R_1 \Big/ \sqrt{\left(1 + \frac{1}{2R_1^2}\right)} \quad . \quad . \quad . \quad (11)$$

$$R_{B2} = \sqrt{2R_1^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

(6.2) Multiplier-Type Detector with Correlated Noise-Signal and Uncorrelated Background Noise

Let the symbols for the uncorrelated noise be the same as in the previous Section, and let ω_p now represent the mid-band angular frequency (approximately).

Let the noise which is correlated between the two inputs be made up of a multiplicity of tones of the type $z \cos (\omega_p - \omega)t$.

Then

$$V_A = z_1 \cos (\omega_p - \omega_1)t + \ldots + z_n \cos (\omega_p - \omega_n)t + x_1 \cos (\omega_p - \omega_{q1})t + \ldots + x_n \cos (\omega_p - \omega_{qn})t \quad . \quad (13)$$

and

$$V_B = z_1 \cos (\omega_p - \omega_1)t + \ldots + z_n \cos (\omega_p - \omega_n)t + y_1 \cos (\omega_p - \omega_{s1})t + \ldots + y_n \cos (\omega_p - \omega_{sn})t \quad . \quad (14)$$

On multiplying these together and removing the h.f. components as before, we obtain

$$\begin{aligned} 2V_A V_B = {} & z_1^2 + z_2^2 + \ldots + z_n^2 \\ & + 2z_1 z_2 \cos (\omega_1 - \omega_2)t + \ldots + 2z_1 z_n \cos (\omega_1 - \omega_n)t \\ & + 2z_2 z_3 \cos (\omega_2 - \omega_3)t + \ldots + 2z_2 z_n \cos (\omega_2 - \omega_n)t \\ & + \; . \\ & \quad . \\ & \quad . \\ & + 2z_{n-1} z_n \cos (\omega_{n-1} - \omega_n)t \end{aligned}$$

($\frac{n(n-1)}{2}$ terms in all)

$$\begin{aligned} & + z_1 y_1 \cos (\omega_1 - \omega_{s1})t + \ldots + z_1 y_n \cos (\omega_1 - \omega_{sn})t \\ & + \; . \\ & \quad . \\ & \quad . \\ & + z_n y_1 \cos (\omega_n - \omega_{s1})t + \ldots + z_n y_n \cos (\omega_n - \omega_{sn})t \end{aligned}$$

(n^2 terms in all)

$+$ similar zx terms (n^2 terms)

$+$ similar xy terms (n^2 terms) . . . . . (15)

If we now make all the z's equal to one another, and also all x's and y's equal, we have in the output

$$\text{Signal} = \tfrac{1}{2}nz^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

(assuming that the correlated noise represents a signal)

$$\text{Noise} = V_{n0} = \tfrac{1}{2}\sqrt{(n^2z^4 + n^2z^2x^2 + \tfrac{1}{2}n^2x^4)} \quad . \quad . \quad (17)$$

where we have assumed that n tends to infinity so that

$$\frac{n(n-1)}{2} \to \frac{n^2}{2}$$

Let V_s be the r.m.s. voltage of the input correlated noise (or "signal"), and V_{n1} be the r.m.s. voltage of the uncorrelated noise on each input. Then

$$R_1 = \frac{V_s}{V_{n1}} = \frac{z}{x}$$

and $V_s = \sqrt{(nz^2/2)}$ and $V_{n1} = \sqrt{(nx^2/2)}$

Thus we may write the output as

$$\text{Signal} = V_s^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

$$\text{Noise} = V_{n0} = V_s^2\sqrt{\left(1 + \frac{1}{R_1^2} + \frac{1}{2R_1^4}\right)} \quad . \quad . \quad (19)$$

and the output signal/noise ratio is

$$R_0 = 1\Big/\sqrt{\left(1 + \frac{1}{R_1^2} + \frac{1}{2R_1^4}\right)} \quad . \quad . \quad . \quad (20)$$

For pulse conditions we obtain

$$R_{A1} = \sqrt{(2)}R_1^2\left[1 + \sqrt{\left(1 + \frac{1}{R_1^2} + \frac{1}{2R_1^4}\right)}\right] \quad . \quad . \quad (21)$$

$$R_{A2} = \sqrt{(2)}R_1^2\sqrt{\left(2 + \frac{1}{R_1^2} + \frac{1}{2R_1^4}\right)} \quad . \quad . \quad . \quad (22)$$

$$R_{B1} = R_0 = 1\Big/\sqrt{\left(1 + \frac{1}{R_1^2} + \frac{1}{2R_1^4}\right)} \quad . \quad . \quad . \quad (23)$$

$$R_{B2} = \sqrt{(2)}R_1^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

(6.3) Single-Rectifier Detector with Noise Signal in Addition to Background Noise

If the noise which is the signal is uncorrelated with the background noise, then in a plain "linear" rectifier the effect of the signal is merely to add to the background on a power basis and so increase the input noise from V_{n1} to $V_{n1}(1 + \Delta)$. If R_1 is, as before, the ratio of signal/noise-background in the input, then

$$1 + \Delta = \sqrt{(1 + R_1^2)} \quad . \quad . \quad . \quad . \quad . \quad (25)$$

Now it is shown by Bennett[5] that

$$\text{D.C. output (due to noise input voltage } V_{n1}) = \sqrt{\left(\frac{1}{2\pi}\right)}V_{n1} \quad . \quad (26)$$

and

$$\text{L.F. noise output (due to noise input voltage } V_{n1}) = \frac{\sqrt{(2 - \pi/2)}}{\pi}V_{n1} \quad . \quad . \quad . \quad . \quad (27)$$

(omitting the constant numerical factor due to the detector efficiency)

so that $\text{D.C. output} \simeq 2(\text{L.F. noise output}) \quad . \quad . \quad (28)$

and both are proportional to the noise input voltage and are therefore proportional to $(1 + \Delta)$.

Thus $$R_{A1} = R_{A2} = 1 + \Delta = \sqrt{(1 + R_1^2)} \quad . \quad . \quad . \quad (29)$$

$$R_{B1} = \frac{2\Delta}{1 + \Delta} = \frac{2\sqrt{(1 + R_1^2)} - 2}{\sqrt{(1 + R_1^2)}} \quad . \quad . \quad . \quad (30)$$

$$R_{B2} = 2\Delta = 2\sqrt{(1 + R_1^2)} - 2 \quad . \quad . \quad . \quad (31)$$

The square-law rectifier can be treated similarly, and gives the result that

$$\text{D.C. output} = \text{L.F. noise output} \quad . \quad . \quad . \quad (32)$$

and both are proportional to the square of the noise input voltage and are therefore proportional to $(1 + R_1^2)$.

Thus $$R_{A1} = R_{A2} = 1 + R_1^2 \quad . \quad . \quad . \quad . \quad . \quad (33)$$

$$R_{B1} = \frac{R_1^2}{1 + R_1^2} \quad . \quad . \quad . \quad . \quad . \quad (34)$$

$$R_{B2} = R_1^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

(6.4) Multiplier-Type Detector with Rectified Coherent Signal and Background Noise

Let the output of one rectifier be

$$V_{O1} = V_O(1 + x_1 \cos \omega_{q1}t + x_2 \cos \omega_{q2}t + \ldots + x_n \cos \omega_{qn}t) \quad . \quad (36)$$

and that of the other rectifier be

$$V_{O2} = V_O(1 + y_1 \cos \omega_{s1}t + y_2 \cos \omega_{s2}t + \ldots + y_n \cos \omega_{sn}t) \quad . \quad (37)$$

Then the product, which is to be unfiltered in any way, is

$$\begin{aligned} V_{O1} \times V_{O2} = V_O^2[&1 + x_1 \cos \omega_{q1}t + x_2 \cos \omega_{q2}t + \ldots + x_n \cos \omega_{qn}t \\ &+ y_1 \cos \omega_{s1}t + y_2 \cos \omega_{s2}t + \ldots + y_n \cos \omega_{sn}t \\ &\left.\begin{array}{l} + \frac{1}{2}x_1y_1 \cos(\omega_{q1} - \omega_{s1})t + \ldots + \frac{1}{2}x_ny_1 \cos(\omega_{qn} - \omega_{s1})t \\ + \frac{1}{2}x_1y_2 \cos(\omega_{q1} - \omega_{s2})t + \ldots + \frac{1}{2}x_ny_2 \cos(\omega_{qn} - \omega_{s2})t \\ + \\ \quad . \\ \quad . \\ + \frac{1}{2}x_1y_n \cos(\omega_{q1} - \omega_{sn})t + \ldots + \frac{1}{2}x_ny_n \cos(\omega_{qn} - \omega_{sn})t \end{array}\right\} n^2 \text{ terms} \\ &+ n^2 \text{ similar terms in } (\omega_q + \omega_s)] \quad . \quad . \quad . \quad . \quad (38) \end{aligned}$$

It should be noted that the d.c. component does not represent the signal alone, but also the mean rectified level of the noise. Thus, to determine values for the various output signal/noise criteria, it is necessary to determine what d.c. and a.c. outputs are produced in the absence of the signal, and what changes in these outputs occur when the signal is applied. Now, omitting a constant factor representing the conversion characteristic of the rectifier in relation to its external circuit, we can quote from existing literature[3] the following formulae for direct-current and r.m.s. alternating-current outputs of linear and square-law rectifiers:

Linear Rectifier.

$$\text{D.C. output} = \frac{V_{n1}}{\sqrt{(2\pi)}}{}_1F_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

$$\text{A.C. output} = \frac{V_{n1}}{\pi}\sqrt{\left[2(1 + R_1^2) - \frac{\pi}{2}({}_1F_1)^2\right]} \quad . \quad (40)$$

where R_1 = Input signal/noise ratio applied to rectifiers

and ${}_1F_1$ is the confluent hypergeometric function[6]

$${}_1F_1(-\tfrac{1}{2}; 1; -R_1^2) \quad . \quad . \quad . \quad . \quad . \quad (41)$$

Square-Law Rectifiers.

$$\text{D.C. output} = V_{n1}^2(1 + R_1^2) \quad . \quad . \quad . \quad . \quad (42)$$

$$\text{A.C. output} = V_{n1}^2\sqrt{(1 + 2R_1^2)} \quad . \quad . \quad . \quad (43)$$

The output when the signal is absent is obviously obtained by putting R_1 equal to zero. In this condition ${}_1F_1 = 1$. Numerical values of ${}_1F_1$ for other values of R_1 are given in printed tables.[6]

Let us now normalize the expressions by making the d.c. output of each rectifier unity when there is no signal, i.e. we put

$$\frac{V_{n1}}{\sqrt{(2\pi)}} = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

for the linear detector, and

$$V_{n1} = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (45)$$

for the square-law detector.

Then in the absence of signal, the a.c. output of each rectifier is

$$\sqrt{\left[\frac{2}{\pi}\left(2 - \frac{\pi}{2}\right)\right]} \simeq 0{\cdot}52 \quad . \quad . \quad . \quad . \quad (46)$$

for the linear detector, and unity for the square-law detector.

If, therefore, all x's and y's are made equal, then in the absence of signal

$$\tfrac{1}{2}nx^2 = \frac{2}{\pi}\left(2 - \frac{\pi}{2}\right) \simeq 0{\cdot}27 \quad . \quad . \quad . \quad . \quad (47)$$

for linear rectifiers, and

$$\tfrac{1}{2}nx^2 = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (48)$$

for square-law rectifiers.

The output of the multiplier is therefore

$$\text{Direct current} = 1 \quad . \quad . \quad . \quad . \quad . \quad (49)$$

R.M.S. of alternating current for the linear rectifiers

$$= \sqrt{\left\{\frac{4}{\pi}\left(2 - \frac{\pi}{2}\right)\left[1 + \frac{1}{\pi}\left(2 - \frac{\pi}{2}\right)\right]\right\}} = \sqrt{\left(\frac{16}{\pi^2} - 1\right)} = 0{\cdot}79$$
$$\quad . \quad . \quad . \quad . \quad (50)$$

and R.M.S. of alternating current

$$\text{for the square-law rectifiers} = \sqrt{3} \quad . \quad (51)$$

Now in the presence of signal, the outputs of the rectifiers are

$$\text{Direct current} = {}_1\mathrm{F}_1 \text{ for linear rectifiers} \quad . \quad . \quad . \quad . \quad (52)$$

or

$$= 1 + R_1^2 \text{ for square-law rectifiers} \quad . \quad (53)$$

$$\text{R.M.S. of alternating current} = \sqrt{\left\{\frac{2}{\pi}\left[2(1 + R_1^2) - \frac{\pi}{2}({}_1\mathrm{F}_1)^2\right]\right\}}$$
$$\quad . \quad . \quad . \quad . \quad (54)$$

for linear rectifiers,

or

$$= \sqrt{(1 + 2R_1^2)} \quad . \quad . \quad . \quad . \quad . \quad (55)$$

for square-law rectifiers.

The output of the multiplier is therefore

$$\text{Direct current} = ({}_1\mathrm{F}_1)^2 \quad . \quad . \quad . \quad . \quad (56)$$

for linear rectifiers,

or

$$(1 + R_1^2)^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (57)$$

for square-law rectifiers.

R.M.S. of alternating current is

$$\sqrt{\left(\frac{4}{\pi}\left\{2(1 + R_1^2) - \frac{\pi}{2}({}_1\mathrm{F}_1)^2\right\}\left\{({}_1\mathrm{F}_1)^2 + \frac{1}{\pi}\left[2(1 + R_1^2) - \frac{\pi}{2}({}_1\mathrm{F}_1)^2\right]\right\}\right)}$$
$$\quad . \quad . \quad . \quad . \quad (58)$$

for linear rectifiers.

or

$$\sqrt{\left\{2(1 + 2R_1^2)\left[(1 + R_1^2)^2 + \tfrac{1}{2}(1 + 2R_1^2)\right]\right\}} \quad . \quad (59)$$

for square-law rectifiers.

We now have all the information we need to calculate R_{A1}, R_{A2}, R_{B1} and R_{B2} in terms of R_1. It is not worth while to write out the full analytic expressions for these criteria, as they are obviously mostly very long. But R_{B2}, which is probably the most useful criterion, is also fairly simple;

thus

$$R_{B2} = \frac{({}_1\mathrm{F}_1)^2 - 1}{0{\cdot}79} \quad . \quad . \quad . \quad . \quad . \quad (60)$$

for linear rectifiers

or

$$\frac{(1 + R_1^2)^2 - 1}{\sqrt{3}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (61)$$

for square-law rectifiers.

All the expressions for the linear-rectifier case simplify very considerably when R_1 is equal to or greater than unity, because then[3]

$${}_1\mathrm{F}_1 \simeq \frac{2}{\sqrt{\pi}} R_1\left(1 + \frac{1}{4R_1^2}\right) \quad . \quad . \quad . \quad . \quad (62)$$

Note that[3] when $R_1 \leqslant 1$,

$${}_1\mathrm{F}_1 \simeq 1 + \tfrac{1}{2}R_1^2 \quad . \quad . \quad . \quad . \quad . \quad (63)$$

so that eqn. (60) becomes

$$R_{B2} \simeq \frac{(1 + \frac{1}{2}R_1^2)^2 - 1}{0{\cdot}79} \simeq \frac{R_1^2}{0{\cdot}79} \quad . \quad . \quad . \quad (64)$$

(6.5) Multiplier-Type Detector with Rectified Noise-Signal and Background Noise

The input to each rectifier consists of noise, part of which is uncorrelated between one input and the other, and part (i.e. the signal) correlated. The inputs can then be expressed as in Section 6.2, eqns. (13) and (14), thus:

$$V_A = z_1 \cos(\omega_p - \omega_1)t + \ldots + z_n \cos(\omega_p - \omega_n)t + x_1 \cos(\omega_p - \omega_{q1})t + \ldots + x_n \cos(\omega_p - \omega_{qn})t \quad . \quad (65)$$

$$V_B = z_1 \cos(\omega_p - \omega_1)t + \ldots + z_n \cos(\omega_p - \omega_n)t + y_1 \cos(\omega_p - \omega_{s1})t + \ldots + y_n \cos(\omega_p - \omega_{sn})t \quad . \quad (66)$$

When the rectifiers are linear, the output of each consists, to a close approximation, of the plus and minus cross-products as discussed in Section 3.3 of Reference 7. Thus the a.c. output of the first rectifier is proportional to

$$\frac{n(n-1)}{2} \text{ terms in all} \left\{ \begin{array}{l} z_1 z_2 \cos(\omega_1 - \omega_2)t + \ldots + z_1 z_n \cos(\omega_1 - \omega_n)t \\ \quad + z_2 z_3 \cos(\omega_2 - \omega_3)t + \ldots + z_2 z_n \cos(\omega_2 - \omega_n)t \\ \quad + \; . \\ \qquad . \\ \qquad . \\ \quad + z_{n-1} z_n \cos(\omega_{n-1} - \omega_n)t \end{array} \right.$$

$$\frac{n(n-1)}{2} \text{ terms} \left\{ \begin{array}{l} \\ \quad - \text{ a similar series in } \omega_1 + \omega_2,\ \omega_2 + \omega_3, \text{ etc.} \\ \\ \end{array} \right.$$

$$\frac{n(n-1)}{2} \text{ terms} \left\{ \begin{array}{l} + x_1 x_2 \cos(\omega_{q1} - \omega_{q2})t + \ldots + x_1 x_n \cos(\omega_{q1} - \omega_{qn})t \\ \quad + \; . \\ \qquad . \\ \qquad . \\ \quad + x_{n-1} x_n \cos[\omega_{q(n-1)} - \omega_{qn}]t \end{array} \right.$$

$$\frac{n(n-1)}{2} \text{ terms} \left\{ \begin{array}{l} \\ \quad - \text{ a similar series in } \omega_{q1} + \omega_{q2},\ \omega_{q2} + \omega_{q3}, \text{ etc.} \\ \\ \end{array} \right.$$

$$n^2 \text{ terms} \left\{ \begin{array}{l} + z_1x_1 \cos(\omega_1 - \omega_{q1})t + \ldots + z_1x_n \cos(\omega_1 - \omega_{qn})t \\ + z_2x_1 \cos(\omega_2 - \omega_{q1})t + \ldots + z_2x_n \cos(\omega_2 - \omega_{qn})t \\ + \ldots \\ + z_nx_1 \cos(\omega_n - \omega_{q1})t + \ldots + z_nx_n \cos(\omega_n - \omega_{qn})t \end{array} \right.$$

$$n^2 \text{ terms} \left\{ \right. + \text{ a similar series in } \omega + \omega_q \quad \ldots \quad (67)$$

and the a.c. output of the second rectifier is similar but with y replacing x and ω_s replacing ω_q.

Now, comparing the a.c. outputs of the two rectifiers, it is clear that the only components which are correlated are those in z and ω exclusively, i.e. the first $n(n-1)$ terms in the above expression. The remaining $n(n-1) + 2n^2$ terms are quite uncorrelated. Putting all the z's equal, and all the x's and y's equal, and letting n tend to infinity, we have

$$R_1 = z/x$$

The signal input V_s is $\sqrt{(nz^2/2)}$.
The noise input V_{n1} is $\sqrt{(nx^2/2)}$.

The d.c. output of each rectifier is $V_{n1}\sqrt{(1 + R_1^2)}/\sqrt{(2\pi)}$. (68)

The correlated a.c. output of each rectifier is

$$\frac{\sqrt{(\frac{1}{2}n^2z^4)}}{\sqrt{(\frac{1}{2}n^2z^4 + \frac{1}{2}n^2x^4 + n^2z^2x^2)}} \frac{V_{n1}\sqrt{(1 + R_1^2)}}{\pi} \sqrt{\left(2 - \frac{\pi}{2}\right)}$$

$$= \frac{0{\cdot}21R_1^2}{\sqrt{(1 + R_1^2)}} V_{n1} \quad . \quad (69)$$

and the uncorrelated a.c. output of each rectifier is

$$0{\cdot}21\sqrt{\left[\frac{(1 + 2R_1^2)}{1 + R_1^2}\right]} V_{n1} \quad \ldots \quad (70)$$

Having determined eqns. (68), (69) and (70) we can now proceed to the multiplying stage. As we do not need to refer further to eqns. (65), (66) and (67), it will be most convenient to use the same symbols over again, but this time z and ω refer to the correlated part of the rectifier outputs, and x, ω_q and y, ω_s refer to the uncorrelated parts of the outputs. Thus, on the lines of Sections 6.2 and 6.4, the inputs to the multiplier are

$$V_{O1} = V_O(1 + z_1 \cos \omega_1 t + z_2 \cos \omega_2 t + \ldots + z_n \cos \omega_n t + x_1 \cos \omega_{q1} t + x_2 \cos \omega_{q2} t + \ldots + x_n \cos \omega_{qn} t) \quad . \quad (71)$$

and

$$V_{O2} = V_O(1 + z_1 \cos \omega_1 t + z_2 \cos \omega_2 t + \ldots + z_n \cos \omega_n t + y_1 \cos \omega_{s1} t + y_2 \cos \omega_{s2} t + \ldots + y_n \cos \omega_{sn} t) \quad . \quad (72)$$

where we shall put all z's equal, and all x's and y's equal, and normalizing as in previous Appendices by putting $V_{n1}/\sqrt{(2\pi)} = 1$, i.e. $V_{n1} = 2{\cdot}5$, we have, from eqns. (68), (69) and (70),

$$V_O = \sqrt{(1 + R_1^2)} \quad \ldots \quad (73)$$

$$V_O\sqrt{(\tfrac{1}{2}nz^2)} = 0{\cdot}525\frac{R_1^2}{\sqrt{(1 + R_1^2)}} \quad \ldots \quad (74)$$

$$V_O\sqrt{(\tfrac{1}{2}nx^2)} = 0{\cdot}525\sqrt{\left(\frac{1 + 2R_1^2}{1 + R_1^2}\right)} \quad \ldots \quad (75)$$

The product is

$$V_{O1} \times V_{O2} = V_O^2\Big[1 + \tfrac{1}{2}z_1^2 + \tfrac{1}{2}z_2^2 + \ldots + \tfrac{1}{2}z_n^2$$

$$n \text{ terms} \left\{ \right. + 2z_1 \cos \omega_1 t + 2z_2 \cos \omega_2 t + \ldots + 2z_n \cos \omega_n t$$

$$n \text{ terms} \left\{ \right. + \tfrac{1}{2}z_1^2 \cos 2\omega_1 t + \tfrac{1}{2}z_2^2 \cos 2\omega_1 t + \ldots + \tfrac{1}{2}z_n^2 \cos 2\omega_n t$$

$$n \text{ terms} \left\{ \right. + x_1 \cos \omega_{q1} t + x_2 \cos \omega_{q2} t + \ldots + x_n \cos \omega_{qn} t$$

$$n \text{ terms} \left\{ \right. + y_1 \cos \omega_{s1} t + y_2 \cos \omega_{s2} t + \ldots + y_n \cos \omega_{sn} t$$

$$\frac{n(n-1)}{2} \text{ terms} \left\{ \begin{array}{l} + z_1z_2 \cos(\omega_1 - \omega_2)t + \ldots + z_1z_n \cos(\omega_1 - \omega_n)t \\ + z_2z_3 \cos(\omega_2 - \omega_3)t + \ldots + z_2z_n \cos(\omega_2 - \omega_n)t \\ + \ldots \\ + z_{n-1}z_n \cos(\omega_{n-1} - \omega_n)t \end{array} \right.$$

$$\frac{n(n-1)}{2} \text{ terms} \left\{ \right. + \text{corresponding terms in } \omega_1 + \omega_2, \text{ etc.}$$

$$n^2 \text{ terms} \left\{ \begin{array}{l} + \tfrac{1}{2}z_1x_1 \cos(\omega_1 - \omega_{q1})t + \ldots + \tfrac{1}{2}z_1x_n \cos(\omega_1 - \omega_{qn})t \\ + \tfrac{1}{2}z_2x_1 \cos(\omega_2 - \omega_{q1})t + \ldots + \tfrac{1}{2}z_nx_n \cos(\omega_2 - \omega_{qn})t \\ + \ldots \\ + \tfrac{1}{2}z_nx_1 \cos(\omega_n - \omega_{q1})t + \ldots + \tfrac{1}{2}z_nx_n \cos(\omega_n - \omega_{qn})t \end{array} \right.$$

$$n^2 \text{ terms} \left\{ \right. + \text{corresponding terms in } (\omega + \omega_q)$$

$$n^2 \text{ terms} \left\{ \right. + \text{corresponding terms of form } \tfrac{1}{2}zy \cos(\omega - \omega_s)t$$

$$n^2 \text{ terms} \left\{ \right. + \text{corresponding terms of form } \tfrac{1}{2}zy \cos(\omega + \omega_s)t$$

$$n^2 \text{ terms} \left\{ \right. + \text{corresponding terms of form } \tfrac{1}{2}xy \cos(\omega_q - \omega_s)t$$

$$n^2 \text{ terms} \left\{ \right. + \text{corresponding terms of form } \tfrac{1}{2}xy \cos(\omega_q + \omega_s)t\Big] \quad \ldots \quad (76)$$

Thus, *in the absence of signal* (i.e. $z = 0$), the output of the multiplier is, on putting all x's and y's equal,

$$\text{Direct current} = 1 \quad \ldots \quad (77)$$

$$\text{R.M.S. of alternating current} = \sqrt{(nx^2 + \tfrac{1}{4}n^2x^4)} = \sqrt{[nx^2(1 + \tfrac{1}{4}nx^2)]} = \sqrt{[0{\cdot}55(1{\cdot}137)]} = 0{\cdot}79 \quad (78)$$

and, *in the presence of signal*, it is:

$$\text{Direct current} = (1 + R_1^2)(1 + \tfrac{1}{2}nz^2)$$

$$1 + R_1^2 + \frac{0{\cdot}275R_1^4}{1 + R_1^2} \quad \ldots \quad (79)$$

R.M.S. of alternating current =

$$(1 + R_1^2)\sqrt{(2nz^2 + nx^2 + \tfrac{1}{2}n^2z^4 + \tfrac{1}{2}n^2z^2x^2 + \tfrac{1}{4}n^2x^4)} \quad . \quad (80)$$

where from eqns. (74) and (75),

$$\left. \begin{array}{l} nz^2 = 0{\cdot}55\dfrac{R_1^4}{(1 + R_1^2)^2} \\ \text{and} \qquad nx^2 = 0{\cdot}55\dfrac{1 + 2R_1^2}{(1 + R_1^2)^2} \end{array} \right\} \quad \ldots \quad (81)$$

Note that in eqn. (80), the terms $z^2 \cos 2\omega t$ in eqn. (76) have been omitted since they are quite negligible.

From the above equations, R_{A1}, R_{A2}, R_{B1} and R_{B2} can be calculated for any value of R_1.

When square-law rectifiers are used the working is similar, except that the output of each rectifier contains only the difference products, no sum products being formed. This makes no difference to the remainder of the working, however; and the only adjustment necessary is that eqns. (68), (69) and (70) are replaced by

$$\text{D.C. output of each rectifier} = V_{n1}^2(1 + R_1^2) \quad . \quad (82)$$

$$\text{Correlated a.c. output of each rectifier} = V_{n1}^2 R_1^2 \quad . \quad . \quad . \quad (83)$$

$$\text{Uncorrelated a.c. output of each rectifier} = V_{n1}^2\sqrt{(1 + 2R_1^2)} \quad . \quad (84)$$

Thus eqns. (73), (74) and (75) are replaced by (on normalizing by putting $V_{n1} = 1$):

$$V_O = 1 + R_1^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (85)$$

$$V_O\sqrt{(\tfrac{1}{2}nz^2)} = R_1^2 \quad . \quad . \quad . \quad . \quad . \quad (86)$$

$$V_O\sqrt{(\tfrac{1}{2}nx^2)} = \sqrt{(1 + 2R_1^2)} \quad . \quad . \quad . \quad . \quad (87)$$

So that the output of the multiplier is, *in the absence of signal*,

$$\left.\begin{aligned}&\text{Direct current} = 1\\&\text{R.M.S. of alternating current} = 1{\cdot}73\end{aligned}\right\} \quad . \quad . \quad (88)$$

and *in the presence of signal*, it is

$$\text{Direct current} = 1 + 2R_1^2 + 2R_1^4 \quad . \quad . \quad . \quad (89)$$

$$\text{R.M.S. of alternating current} =$$
$$(1 + R_1^2)^2\sqrt{(2nz^2 + nx^2 + \tfrac{1}{2}n^2z^4 + \tfrac{1}{2}n^2z^2x^2 + \tfrac{1}{4}n^2x^4)} \quad . \quad (90)$$

where from eqns. (86) and (87),

$$\left.\begin{aligned}nz^2 &= \frac{2R_1^4}{(1 + R_1^2)^2}\\ \text{and}\qquad nx^2 &= \frac{2(1 + 2R_1^2)}{(1 - R_1^2)^2}\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (91)$$

(6.6) Addition-Type Detector with Rectified Coherent Signal and Background Noise (otherwise called "Pulse-to-Pulse Integration")

Let the output of one rectifier be

$$V_{O1} = V_O(1 + x_1\cos\omega_{q1}t + x_2\cos\omega_{q2}t + \ldots + x_n\cos\omega_{qn}t) \quad . \quad (92)$$

and that of the other rectifier be

$$V_{O2} = V_O(1 + y_1\cos\omega_{s1}t + y_2\cos\omega_{s2}t + \ldots + y_n\cos\omega_{sn}t) \quad . \quad (93)$$

Then the sum, which is to be unfiltered, is

$$V_{O1} + V_{O2} = V_O[2 + x_1\cos\omega_{q1}t + \ldots + x_n\cos\omega_{qn}t - y_1\cos\omega_{s1}t - \ldots + y_n\cos\omega_{sn}t] \quad . \quad (94)$$

Now the d.c. components in eqns. (92) and (93) represent partly the signal and partly the mean rectified level of the noise. The appropriate formulae relating the output of signal and noise from each rectifier to the input signal/noise ratio R_1 are given in Section 6.4. Thus, considering both linear and square-law rectifiers, we obtain by the use of eqns. (46), (52) to (55) and (94), on putting all x's and y's equal, the following relations:

In the absence of signal, the output of the adder is

$$\text{Direct current} = 2 \quad . \quad . \quad . \quad . \quad . \quad (95)$$

$$\text{R.M.S. of alternating current} = \sqrt{\left[\frac{4}{\pi}\left(2 - \frac{\pi}{2}\right)\right]} = 0{\cdot}74 \quad . \quad (96)$$

for linear rectifiers,

$$\text{R.M.S. of alternating current} = \sqrt{2} \quad . \quad . \quad (97)$$

for square-law rectifiers.

In the presence of signal, the output of the adder is

$$\text{Direct current} = 2({}_1F_1) \quad . \quad . \quad . \quad . \quad (98)$$

for linear rectifiers,

$$\text{or} \qquad = 2(1 + R_1^2) \quad . \quad . \quad . \quad . \quad . \quad (99)$$

for square-law rectifiers.

$$\text{R.M.S. of alternating current} = \sqrt{\left\{\frac{4}{\pi}\left[2(1 + R_1^2) - \frac{\pi}{2}({}_1F_1)^2\right]\right\}}$$

for linear rectifiers,

$$\text{or} \qquad = \sqrt{[2(1 + 2R_1^2)]} \quad . \quad . \quad . \quad . \quad (101)$$

for square-law rectifiers.

From this information, R_{A1}, R_{A2}, R_{B1} and R_{B2} can all be calculated. In particular,

$$R_{B2} = \frac{2[{}_1F_1 - 1]}{\sqrt{\left[\frac{4}{\pi}\left(2 - \frac{\pi}{2}\right)\right]}} = \frac{{}_1F_1 - 1}{0{\cdot}37} \quad . \quad . \quad (102)$$

for linear rectifiers,

$$\text{or} \qquad = \sqrt{(2)}R_1^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (103)$$

for square-law rectifiers.

It should be noted that, whatever the law of the detector, R_{B1} and R_{B2} are always $\sqrt{2}$ times the value obtained for a single rectifier with the same input signal/noise ratio. R_{A1} and R_{A2} are more complex, but are always less than $\sqrt{2}$ times the single-rectifier value.

(6.7) Addition-Type Detector with Rectified Noise-Signal and Background Noise

The working combines the principles of the first part of Section 6.5 and of Section 6.6.

Section 6.4 shows that the output of the two rectifiers comprises direct current, correlated alternating current and uncorrelated alternating current, with the following normalized values:

$$\text{Linear rectifiers: Direct current} = \sqrt{(1 + R_1^2)} \quad . \quad (104)$$

$$\text{Correlated alternating current} = 0{\cdot}525\frac{R_1^2}{\sqrt{(1 + R_1^2)}} \quad . \quad (105)$$

$$\text{Uncorrelated alternating current} = 0{\cdot}525\sqrt{\left(\frac{1 + 2R_1^2}{1 + R_1^2}\right)} \quad . \quad (106)$$

$$\text{Square-law rectifiers: Direct current} = 1 + R_1^2 \quad . \quad (107)$$

$$\text{Correlated alternating current} = R_1^2 \quad . \quad . \quad (108)$$

$$\text{Uncorrected alternating current} = \sqrt{(1 + 2R_1^2)} \quad . \quad (109)$$

It is clear that on addition of the outputs of the two rectifiers, we obtain:

$$\text{For linear rectifiers, direct current} = 2\sqrt{(1 + R_1^2)} \quad . \quad (110)$$

$$\text{and alternating current} = 1{\cdot}05\frac{R_1^2}{\sqrt{(1 + R_1^2)}} + 0{\cdot}74\sqrt{\left(\frac{1 + 2R_1^2}{1 + R_1^2}\right)} \quad . \quad . \quad . \quad . \quad (111)$$

$$\text{For square-law rectifiers, direct current} = 2(1 + R_1^2) \quad . \quad (112)$$

$$\text{and alternating current} = 2R_1^2 + \sqrt{[2(1 + 2R_1^2)]} \quad . \quad . \quad (113)$$

The outputs in the absence of signal are obtained by putting R_1 equal to zero. From this information, R_{A1}, R_{A2}, R_{B1} and R_{B2} can readily be calculated.

621.396.96: 621–526: 621.3.016.35

Monograph No. 121 R
Mar. 1955

THE STABILITY AND TIME RESPONSE OF FAST-OPERATING CLOSED-LOOP PULSED RADAR CIRCUITS

By D. McDONNELL and W. R. PERKINS, B.Sc.

(*The paper was first received* 27*th December*, 1953, *in revised form* 26*th May and in final form* 25*th November*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in March*, 1955.)

SUMMARY

The paper deals with the stability and time responses of a sampling servo-system typical of a.g.c., a.f.c., range-measuring and overall-feedback circuits used in pulsed radar equipments in which the loop-response time is not many times greater than the pulse-repetition time. The case of a high-speed radar-controlled missile may be such a system.

Conditions under which Nyquist plots can be made are given. Expressions for the output both at and between the sampling times are obtained in terms of the input functions of the system. The operation of the systems in the presence of noise is considered in the Appendix. A short list of transformations suitable for analysis is included.

(1) INTRODUCTION

In many pulsed radar equipments a.g.c. and a.f.c. circuits are used to ensure that a suitable video signal is available. The video signal may operate automatic range-measuring circuits or may be resolved to give space co-ordinates. The three units—i.e. a.g.c., a.f.c. and range-measuring units—are generally closed-loop systems. In the case of the video signal being resolved to give space co-ordinates, a closed loop is sometimes formed via an aerodynamic link or other device with two or three degrees of freedom. In all four cases mentioned the loop stability and time response may be influenced considerably by the sampled nature of the basic radar information.

Several theoretical methods have been given by which sampling servo-systems may be analysed, but the author has not found any paper giving formulae in forms suitable for easy use by the engineer. In the present paper a transform method is used, and characteristic equations involving the transform are given for several typical closed-loop circuits encountered in radar equipments. A characteristic equation is obtained for a very general system. The output-time function corresponding to a particular input function can be obtained by use of the characteristic equation and a convolution sum. When the characteristic equation contains polynomials of high degree, Nyquist plots can be made in a manner similar to that for the continuous case. Expressions for the output between the sampling times are obtained.

Before the analysis of the different loops is shown, some discussion of the mathematical tools used is necessary, and this is given in Section 2. In Section 3 a characteristic equation for a general sampling system is given. In Section 4 practical systems are considered. Section 5 concludes with some examples on the use of the theory in typical systems.

(2) TRANSFORM PAIRS

The Taylor expansion of a function of a complex variable forms the basic transform used in the paper. Eqns. (1) and (2) and the conditions of validity can be obtained from any treatise on functions of a complex variable.

If $G(z)$ is a function of the complex variable z under certain conditions, it may be expanded about the point $z = 0$:

$$G(z) = \sum_{n=0}^{n=\infty} A_n z^n \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The coefficients A_r are given by

$$A_r = G^r(0)/r! \; [G^r(0) \text{ is the } r\text{th derivative of } G(z) \text{ at } z = 0]$$

Also
$$G^r(0) = \frac{r!}{2\pi i}\int_{c'} \frac{G(z)}{z^{r+1}} dz$$

The contour c' is a small circle around the origin. Substitution gives

$$A_r = \frac{1}{2\pi i}\int_{c'} \frac{G(z)}{z^{r+1}} dz \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

It is convenient to replace A_r by a function of r which is equal to it for r integral;

i.e. let $A_r = f(r)$ for r an integer.

Making this substitution and also putting $z = 1/q$ and $G(1/q) = \mathscr{F}(q)$ in eqns. (1) and (2) gives

$$\mathscr{F}(q) = \sum_{n=0}^{n=\infty} f(n) q^{-n} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1a)$$

$$f(n) = \frac{1}{2\pi i}\int_{c} \mathscr{F}(q) q^{n-1} dq \quad . \quad . \quad . \quad . \quad . \quad (2a)$$

c' is replaced by c, which is a large circle around the origin.

Eqns. (1*a*) and (2*a*) form a transform pair. $\mathscr{F}(q)$ is referred to as the discontinuous transform of the function $f(n)$, and $f(n)$ as the inverse discontinuous transforms of $\mathscr{F}(q)$.

The discontinuous transform can be used to transform linear difference and summation equations into algebraic form in a way similar to that in which the Laplace transformation is used for differential equations. It is, in fact, possible to deduce the Fourier integral theorem and the Laplace transformation as being the limiting cases of eqns. (1*a*) and (2*a*). It is not the purpose of the present paper to show that this is so, but the reader will find the analysis which follows much easier to understand if the correspondence between this finite-difference case and the limiting case of Fourier or Laplace is noted. Shifting and convolution theorems similar to those encountered in Laplace transformations are included in the list of transforms given in Table 1.

If $F(p)$ is the Laplace transform of $f(t)$ then $\mathscr{F}(q)$ and $F(p)$ form a further transform pair:

$$\mathscr{F}(q) = \frac{1}{2\pi i}\int_{Br} F(p) \frac{q}{q - \varepsilon^p} dp \quad . \quad . \quad . \quad . \quad (3)$$

$$F(p) = \frac{1}{2\pi i}\int_{c} \mathscr{F}(q) \frac{1}{q(p - \log_\varepsilon q)} dq \quad . \quad . \quad . \quad (4)$$

Br is a Bromwich contour and c is a large circle around the origin.

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. McDonnell and Mr. Perkins are with Vickers-Armstrongs, Ltd.

Table 1

$f(n)$	$\sum_{n=0}^{n=\infty} f(n)q^{-n} \equiv \mathscr{D}[f(n)]$
$f(n-m)$	$q^{-m}\sum_{n=0}^{n=\infty} q^{-n}f(n) + q^{-m}\sum_{n=-1}^{n=-m} q^{-n}f(n)$ (The second term is often zero)
$\sum_{r=0}^{r=n} f(n-r)g(r)$	$\sum_{n=0}^{n=\infty} q^{-n}f(n) \sum_{n=0}^{n=\infty} q^{-n}g(n)$
$\Delta^s V(n)$ [Sth difference of $V(n)$]	$(q-1)^s \sum_{n=0}^{n=\infty} q^{-n}V(n) - q[\Delta^{s-1}V(0) + (q-1)\Delta^{s-2} V(0) + (q-1)^2\Delta^{s-3}V(0) \ldots (q-1)^{s-1}V(0)]$
$\left[\sum_{r=0}^{r=n}\right]^s V(r)$*	$\frac{q^s}{(q-1)^s}\sum_{n=0}^{n=\infty} q^{-n}V(n) + \left\{\frac{q}{(q-1)}\left[\sum_0\right]^s V(r) + \frac{q^2}{(q-1)^2}\left[\sum_0\right]^{s-1} V(r) \ldots + \frac{q^s}{(q-1)^s}\sum_0 V(r)\right\}$
K-constant	$\frac{qK}{q-1}$
$n^{(m)} = n(n-1)\ldots(n-m+1)$	$\frac{m!q}{(q-1)^{m+1}}$
a^n	$\frac{q}{q-a}$
$\sin an$	$\frac{q \sin a}{q^2 - 2q\cos a + 1}$
$\cos an$	$\frac{q(q-\cos a)}{q^2 - 2q\cos a + 1}$
Pulse of amplitude A at time n	$\frac{A}{q^n}$

* The symbol $\left[\sum_{r=0}^{r=n}\right]^s V(r)$ is equal to:

$$\sum_{r_s=0}^{r_s=n}\sum_{r_{s-1}=0}^{r_{s-1}=r_s}\cdots\sum_{r=0}^{r=r_1} V(r)$$

On the right-hand-side symbol $\left[\sum_0\right]^s$ is the same as above except that the lower limit only is taken in the last summation.

Eqns. (3) and (4) are obtained by substitution of the Laplace transform and changing the order of summation or integration. Eqn. (3) is sometimes useful for obtaining $\mathscr{F}(q)$ direct from $F(p)$ instead of obtaining $f(t)$ by the inverse Laplace transform and then using eqn. (1*a*).

In the analysis which follows, expressions containing infinite and finite sums are encountered, but by suitable use of the transform pair 1(*a*) and 2(*a*) and the transform of the convolution sum, together with the shifting theorem, manipulation becomes very simple.

(3) GENERAL SAMPLING SYSTEM

Fig. 1 shows a typical sampling servo-system which is split into three parts, namely (*a*) the sampler, (*b*) the discontinuous shaping circuits and (*c*) the passive shaping circuit.

(3.1) Sampler

The sample unit samples the difference between the input and the output and maintains the sampled value until the next sampling time. (The intervals between the sampling times are equal and normalized to unity.) The output from the sampling unit between the sampling times is equal to the value of the last sample only and is not dependent upon previous samples.

(3.2) Discontinuous Shaping Circuits

The discontinuous shaping circuits modify the output from the sampling unit by any linear combination of summing and differencing. The output from this unit is constant between the sampling times. Suppose, for example, that the output from the sampler is $v_s(n)$; the discontinuous shaping unit may operate on $v_s(n)$ to produce an output v_D given by

$$v_D = \sum^{r_1} \left\{ \sum_0^n \right\}^{r_1} v_s(r) + \sum^{s_1} \Delta^{s_1} v_s(n) \quad . \quad . \quad . \quad (6)$$

i.e. v_D may be formed by terms like the r_1th sum added to terms like the s_1th difference of $v_s(n)$.

Another operation which this unit may perform is to delay the information by an integral number of sampling periods, in which case eqn. (6) becomes

$$v_D = \sum^{r_1} \left\{ \sum_0^n \right\}^{r_1} v_s(r - m_{r_1}) + \sum^{s_1} \Delta^{s_1} v_s(r - m_{s_1}) \quad . \quad (6a)$$

Exponentially weighted sums, etc., can also be included.

(3.3) Passive Shaping Circuit

The passive shaping circuit contains inductive, capacitive and resistive elements, and has a step response given by the inverse Laplace transform of F(p), i.e. f(t).

For the system of Fig. 1 the output at the nth sampling time is given by

$$v_o(0) = v_D(0)\mathrm{f}(0)$$
$$v_o(1) = v_D(0)\mathrm{f}(1) + [v_D(1) - v_D(0)]\mathrm{f}(0)$$
$$v_o(2) = v_D(0)\mathrm{f}(2) + [v_D(1) - v_D(0)]\mathrm{f}(1) + [v_D(2) - v_D(1)]\mathrm{f}(0)$$
$$v_o(2) = v_D(0)[\mathrm{f}(2) - \mathrm{f}(1)] + v_D(1)[\mathrm{f}(1) - \mathrm{f}(0)] + v_D(2)\mathrm{f}(0)$$

etc.

$$v_o(n) = \sum_{r=0}^{r=n} v_D(r)\Delta\mathrm{f}(n - r - 1)$$

Taking discontinuous transforms of both sides gives

$$\mathscr{D}[v_o(n)] = \mathscr{D}[v_D(n)]\mathscr{D}[\mathrm{f}(n)]\frac{(q-1)}{q}$$

But from eqn. (6a), and using the table of transforms

$$\mathscr{D}[v_D(n)] = \left[\sum^r \frac{q^r}{q^{mr}(q-1)^r} + \sum^s \frac{(q-1)^s}{q^{m_s}}\right]\mathscr{D}[v_s(n)]$$
$$= \phi(q)\mathscr{D}[v_s(n)]$$

Therefore $$\mathscr{D}[v_o(n)] = \mathscr{D}[v_s(n)]\mathscr{D}[\mathrm{f}(n)]\phi(q)\frac{(q-1)}{q}$$

Substituting $v_1(n) - v_o(n)$ for $v_s(n)$ gives

$$\mathscr{D}[v_o(n)] = \left\{\mathscr{D}[v_1(n)] - \mathscr{D}[v_o(n)]\right\}\mathscr{D}[\mathrm{f}(n)]\frac{\phi(q)(q-1)}{q}$$

or $$\mathscr{D}[v_o(n)] = \mathscr{D}[v_1(n)]\frac{\mathscr{D}[\mathrm{f}(n)]\phi(q)(q-1)}{q + \mathscr{D}[\mathrm{f}(n)]\phi(q)(q-1)} = \theta(q)\mathscr{D}[v_1(n)]$$
$$. \quad . \quad . \quad . \quad (7)$$

Using the inversion integral gives

$$v_o(n) = \frac{1}{2\pi i}\int_c \mathscr{D}[v_1(n)]\theta(q)q^{n-1}dq \quad . \quad . \quad . \quad (8)$$

$\theta(q)$ is the characteristic equation for the system of Fig. 1.

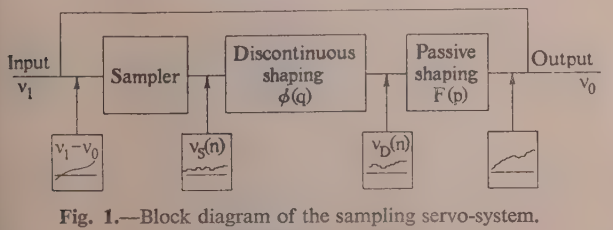


Fig. 1.—Block diagram of the sampling servo-system.
Waveforms are only symbolic.

The stability of the system can be investigated by plotting $\theta(q)$ as a function of q or by observing the position of the roots of the denominator of $\theta(q)$. Before discussing the stability conditions, expressions for the output between the sampling times will be obtained.

By considering the output voltage at a time $(n + \delta t)$ the output between the sampling times can be obtained, $(\delta t < 1)$ for

$$v_o(\delta t) = v_D(0)\mathrm{f}(\delta t)$$
$$v_o(1 + \delta t) = v_D(0)\mathrm{f}(1 + \delta t) + [v_D\mathrm{f}(1) - v_D(0)]\mathrm{f}(0 + \delta t)$$
$$v_o(2 + \delta t) = v_D(0)\mathrm{f}(2 + \delta t) + [v_D(1) - v_D(0)]\mathrm{f}(1 + \delta t)$$
$$+ [v_D(2) - v_D(1)]\mathrm{f}(\delta t)$$
$$v_o(n + \delta t) = \sum_{r=0}^{r=n} v_D(r)\Delta\mathrm{f}(n - r + \delta t - 1)$$

Taking discontinuous transforms of both sides gives

$$\mathscr{D}[v_o(n + \delta t)] = \mathscr{D}[v_D(n)]\mathscr{D}[\mathrm{f}(n + \delta t)]\frac{(q-1)}{q}$$

Substitution for $v_D(n)$ as before gives

$$\mathscr{D}[v_o(n + \delta t)] = \phi(q)\mathscr{D}[v_s(n)]\mathscr{D}[\mathrm{f}(n + \delta t)]\frac{(q-1)}{q}$$

Substitution for $v_s(n)$ gives

$$\mathscr{D}[v_o(n + \delta t)] = \left\{\mathscr{D}[v_1(n)] - \mathscr{D}[v_o(n)]\right\}[\mathrm{f}(n + \delta t)]\frac{(q-1)}{q}\phi(q)$$

$$\frac{\mathscr{D}[v_o(n + \delta t)]}{\mathscr{D}[v_1(n)]} = \frac{(q-1)}{q}\phi(q)\left\{1 - \frac{\mathscr{D}[v_o(n)]}{\mathscr{D}[v_1(n)]}\right\}\mathscr{D}[\mathrm{f}(n + \delta t)]$$

by substitution from eqn. (7)

$$= \frac{(q-1)}{q}\phi(q)\left\{1 - \frac{\mathscr{D}[\mathrm{f}(n)]\phi(q)(q-1)}{q + \mathscr{D}[\mathrm{f}(n)]\phi(q)(q-1)}\right\}\mathscr{D}[\mathrm{f}(n + \delta t)]$$
$$= \frac{(q-1)\phi(q)\mathscr{D}[\mathrm{f}(n + \delta t)]}{q + \phi(q)\mathscr{D}[\mathrm{f}(n)](q-1)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9a)$$

For F(p), f(t) may often be considered to be the finite sum of exponential terms, i.e.

$$\mathrm{f}(t) = \sum^r B_r \varepsilon^{\alpha_r t}$$

in which case

$$\mathscr{D}[\mathrm{f}(n)] = \sum^r B_r \frac{q}{q - \varepsilon^{\alpha_r}}$$

and $$\mathscr{D}[\mathrm{f}(n + \delta t)] = \sum^r B_r \varepsilon^{\alpha\,\delta t}\frac{q}{q - \varepsilon^{\alpha_r}}$$

Thus

$$v_o(n + \delta t) = \frac{1}{2\pi i}\int_c \theta(q)\frac{\sum^r B_r\varepsilon^{\alpha_r\delta t}/(q - \varepsilon^{\alpha_r})}{\sum^r B_r/(q - \varepsilon^{\alpha_r})}\mathscr{D}[v_1(n)]q^{n-1}dq \quad . \quad (9)$$

Eqn. (9) seems to be the general expression for the output either at or between the sampling times for sampling servo-systems of the type shown in Fig. 1, when F(p) contains only simple poles.

(3.4) Nyquist Plots and Conditions for Stability

From eqn. (8) it is clear that the integral converges when the poles of $\theta(q)$ lie within the unit circle centred at the origin; for stability, therefore, the roots of the denominator of $\theta(q)$ must have moduli less than unity. If $\theta(q)$ is such that factorization is not convenient Nyquist plots can be made to determine the stability.

From the principle of the argument in complex-variable theory we have that the number of poles within a contour is given by

$$\text{Number of poles} = -\frac{1}{2\pi i}[\text{Variation along } c \text{ of argument of } \mathrm{f}(z)]$$

It follows that the stability may be checked by taking q around the unit circle and noting whether $\theta(q)$ encircles the origin, in

which case the system is unstable. (The substitution of $z = 1/q$ inverts the enclosed area and the exterior area.)

A practical method of checking the stability can be found when it is possible to expand $\theta(q)$ about the point $q = 1$ in the form

$$\theta(q) = \sum_{n=0}^{n=r_1} A(n)(q-1)^n + \sum_{n=0}^{n=r_2} B(n)q^n/(q-1)^n$$

This can be considered to have originated from a summation-difference equation, Fig. 2, in which the box is equivalent

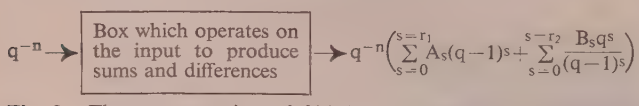


Fig. 2.—The representation of $\theta(q)$ by a summing and differencing device.

to the system. If q^{-n} is applied to the system and q is taken around the unit circle, the plot of output against input gives the equivalent Nyquist plot.

If q^{-n} is written as $\varepsilon^{j\omega n}$, ω must be taken from 0 to 2π. It is also clear from simple practical considerations that it is not necessary to increase the frequency above the sampling frequency.

(4) PRACTICAL SYSTEMS

Many practical a.g.c., a.f.c. and the simpler range-measuring circuits have diodes or non-linear elements connected in such a way that they cannot truly be considered as linear discontinuous devices; particular examples of this are the common diode rectifier used in a.g.c. systems in which the "rising" time-constant is shorter than the "following" time-constant, and the F. C. Williams early-and-late strobe-rectifier system in which a backlash region is allowed before the diodes can conduct. The recent tendency, however, seems to be towards the clamped gate and summing systems in which the operation conforms more exactly with the mathematical process.

If the non-linear nature of some of the time-constants is overlooked, most of the elementary a.g.c. and range-measuring circuits commonly encountered fall into two groups, namely clamped-gate types and summing types. These groups can again be divided into those which delay the information (one pulse late) and those which work instantaneously.

The following discussion deals with typical a.g.c. and range-measuring units from the standpoint of the sampling servo-system. $\theta(q)$'s are worked out for the various systems and discontinuous sampling circuits.

(4.1) Range-measuring Circuits

A type of range measuring circuit for which the sampling servo approach is most suited is the early-and-late strobe-type of device when worked at low pulse repetition frequencies. Two types of unit in common use have discontinuous shaping circuits with q transforms:

$\phi(q) = 1$ for the clamped-gate type.

$\phi(q) = q/(q-1)$ for the summing-gate type (obtained from Table). The relevant $\theta(q)$ values are

$$\theta_s(q) \text{ for the summing gate} = \frac{\mathscr{Z}[\mathrm{f}(n)]}{1 + \mathscr{Z}[\mathrm{f}(n)]} \quad . \quad (10)$$

$$\theta_c(q) \text{ for the clamped gate} = \frac{(q-1)\mathscr{Z}[\mathrm{f}(n)]}{q + (q-1)\mathscr{Z}[\mathrm{f}(n)]} \quad . \quad (11)$$

These are obtained direct by use of Fig. 1 and eqn. (7). The output between the sampling times can be obtained by use of eqn. (9).

In some circuits condensers are used to store the sampled information between the sampling times, leakage of charge on this condenser sometimes being a significant factor. By including in $\mathrm{F}(p)$ a term $pT/(1 + pT)$, where T is the leakage time-constant, this effect is included in the theory.

Use of the inversion integral.—$\mathrm{F}(p)$ can usually be expressed as

$$\mathrm{F}(p) = \sum^{r} \frac{A_r}{(p - \alpha_r)}$$

and therefore $$\mathrm{f}(n) = \sum^{r} A_r \varepsilon^{\alpha_r n}$$

Taking discontinuous transforms gives

$$\mathscr{Z}[\mathrm{f}(n)] = \sum^{r} A_r \frac{q}{q - \varepsilon^{\alpha_r}}$$

The functions of q, $\theta(q)$ are then of the type

$$\mathscr{Z}[v_1(n)]\theta(q) = \frac{\sum^{r} A_r \dfrac{q}{q - \varepsilon^{\alpha_r}} \phi(q)(q-1)\mathscr{Z}[v_1(n)]}{q + \sum^{r} A_r \dfrac{q}{q - \varepsilon^{\alpha_r}} \phi(q)(q-1)}$$

The function $\theta(q)$ must now be expanded as a series in q or as partial fractions;

if $$\mathscr{Z}[v_1(q)]\theta(q) = \sum^{s} \frac{B_s}{(q - \beta_s)}$$

then $$\frac{1}{2\pi i}\int_c \mathscr{Z}[v_1(n)]\theta(q) q^{n-1} dq$$

is given by $$v_0(n) = \sum^{s} B_s \beta_s^{n-1}$$

Multiple Poles.—When $\mathrm{F}(p)$ contains multiple poles, its inverse transform, i.e. $\mathrm{f}(t)$, contains terms like

$$\mathrm{f}(t) = \sum^{r} t^{s_r} \varepsilon^{-\alpha_r t}$$

The discontinuous transform of $\mathrm{f}(n)$ is therefore

$$\mathscr{Z}[\mathrm{f}(n)] = \sum^{r} \sum_{n=0}^{n=\infty} (q\varepsilon^{\alpha_r})^{-n} n^{s_r}$$

Writing $q\varepsilon^{\alpha_r}$ as q'_r

$$[\mathrm{f}(n)] = \sum^{r} \sum_{n=0}^{n=\infty} q'_r{}^{-n} n^{s_r}$$

The problem is then to find the transform of n^{s_r} for s_r up to, say, 4. The list of transforms contains the transform of $n^{(m)}$. By expressing n^s in terms of $n^{(s)}$, $n^{(s-1)}$, etc., the transform of n^s can be found.

If $\theta(q)$ contains multiple poles, use of the inverse Taylor formula can be made by substituting for $\mathrm{G}(z)$:

$$\frac{\mathrm{G}^r(a)}{r!} = \frac{1}{2\pi i}\int_c \frac{\mathrm{G}(z)}{(z-a)^{r+1}} dz$$

(4.1.1) I.F. Amplifier Gain Characteristics.

The amplifier considered has uncontrolled stages of gain G_0, and r stages whose gain is controllable by changing the mutual conductance of the amplifying valves. Fig. 3 shows the arrangement.

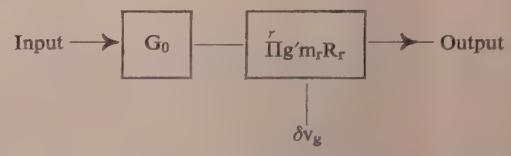


Fig. 3

The gain of the controlled stages is given by the product of the individual gains, each of which is taken, at mid-band, to be gmR, where R is the damping resistance. The overall gain of the amplifier is given by

$$\text{Gain} = G_0 \overset{r}{\Pi} g'm_r R_r$$

If $g'm$ is written as

$$g'm = gm_r + \delta gm_r$$

the gain is given by

$$\text{Gain} = G_0 \overset{r}{\Pi} (gm_r + \delta gm_r)R_r$$

Writing the product out fully and regrouping terms gives

$$\text{Gain} = G_0\left(\overset{r}{\Pi} gm_r R_r + \sum_r \frac{\delta gm_r}{gm_r} \overset{r}{\Pi} gm_r R_r\right)$$

Writing $G = \overset{r}{\Pi} gm_r R_r$

$$\text{Gain} = G_0 G\left(1 + \sum^r \frac{\delta gm_r}{gm_r}\right)$$

$$= G_0 G\left(1 + \delta v_g \sum^r \frac{\delta \log gm_r}{\delta v_g}\right)$$

Assuming that $\delta \log gm_r/\delta v_g = m_r$ for a particular value is constant, we have

$$\text{Gain} = G_0 G\left(1 + \delta v_g \sum^r m_r\right)$$

$$\text{Gain} = G_0 G(1 + \delta v_g K) \quad . \quad . \quad . \quad . \quad . \quad (12)$$

$$K = \sum^r m_r$$

Expression (12) gives the gain when a bias voltage is applied to the a.g.c. line; this bias voltage will be referred to as $v_B(t)$ as a function of time or $v_B(n)$ for discrete times.

(4.1.2) General Arrangement of the System.

The general arrangement under consideration is shown in Fig. 4.

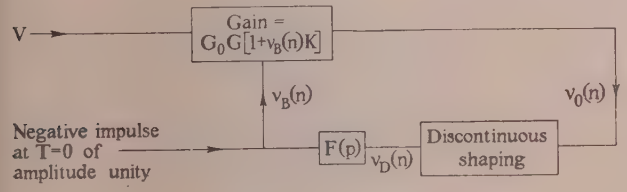


Fig. 4.—Block diagram of a.g.c. system.

The input to the i.f. amplifier, V, is in the form of pulses at times $t = 0$, $t = 1$, $t = 2 \ldots t = n$. The pulse repetition frequency has been normalized to unity and therefore appropriate modifications must be made to $F(p)$ when considering pulse repetition frequencies other than unity. The output of the amplifier $v_0(n)$ is in the form of pulses at times $t = \Delta t$, $t = 1 + \Delta t \ldots t = n + \Delta t$, where Δt is the delay in the amplifier; this delay, however, will be neglected throughout this note. The discontinuous shaping unit changes the pulse information into steps; several types are possible, and the three types considered are listed on the next paragraph.

The unit shown with a step response as the inverse Laplace transform of $F(p)$ is the smoothing network which may contain twin T-filters, etc. Since we are concerned with the stability of the system some method of "shocking" the circuit must be provided; for this purpose an impulse is applied to the gain line as shown on the diagram, Fig. 4.

(4.1.3) Discontinuous Shaping.

The three arrangements for the pulse/direct-current convertor considered are listed below:

Arrangement (i).

$$v_D(n) = v_0(n - 1)$$

In this case the output $v_D(n)$ is in the form of steps, and the amplitude of the step is equal to the previous pulse amplitude. The step fills the gap between the pulses; i.e. the output $v_D(n)$ from $t = n$ to $t = n + 1$ is equal to the pulse amplitude $v_0(n - 1)$.

Arrangement (ii).

$$v_D(n) = v_0(n)$$

In this case the output $v_D(n)$ from time $t = n$ to $t = n + 1$ is equal to the pulse amplitude $v_0(n)$ at time $t = n$.

Arrangement (iii).

$$v_D(n) = \sum_{\psi=0}^{\psi=n-1} v_0(n - \psi - 1)\varepsilon^{-\alpha\psi}$$

The output $v_D(n)$ is due to the exponentially weighted sum of the past values of $v_0(n)$, except the last.

(4.1.4) Mathematical Treatment of the Complete Loop.

The voltage $v_B(n)$ in terms of $v_D(n)$ is given by

$$v_B(n) = \sum_{r=0}^{r=n} v_D(n - r)\mathrm{f}(r) \text{ minus an impulse at } t = 0$$

The voltage $v_B(n)$ in terms of $v_0(n)$ is different for the three convertor cases, namely

Case (i):

$$v_B(n) = \sum_{r=0}^{r=n} v_0(n - r - 1)\mathrm{f}(r) \text{ minus an impulse at } t = 0$$

Case (ii):

$$v_B(n) = \sum_{r=0}^{r=n} v_0(n - r)\mathrm{f}(r) \text{ minus an impulse at } t = 0$$

Case (iii):

$$v_B(n) = \sum_{r=0}^{r=n} \mathrm{f}(r) \sum_{\psi=0}^{\psi=n-1} v_0(n-r-\psi-1)\varepsilon^{-\alpha\psi} \text{ minus an impulse at } t=0.$$

Using eqn. (12) and substituting $v_B(n)$ for case (i) we have

$$v_0(n) = VG_0G \left[1 + K \sum_{r=0}^{r=n} v_0(n-r-1)\mathrm{f}(r) \text{ minus an impulse at } t=0\right]$$

Taking discontinuous transforms of both sides,

$$\mathscr{D}[v_0(n)] = \frac{VG_0G}{q-1}q + \frac{VG_0GK}{q}\left\{\mathscr{D}[v_0(n)]\mathscr{D}[\mathrm{f}(n)]\right\} - VG_0G$$

$$\mathscr{D}[v_0(n)] = \frac{VG_0Gq}{(q-1)(q - VG_0GK\mathscr{D}[\mathrm{f}(n)]} = \theta_1(q) \quad . \quad (13)$$

Using eqn. (12) and substituting for case (2), we have

$$v_0(n) = VG_0G\left[1 + K\sum_{r=0}^{r=n} v_0(n-r)\mathrm{f}(r) - \text{impulse at } t = 0\right]$$

Taking discontinuous transforms of both sides gives

$$\mathscr{D}[v_0(n)] = \frac{VG_0G}{q-1}q + KVG_0G\mathscr{D}[v_0(n)]\mathscr{D}[\mathrm{f}(n)] - VG_0G$$

$$\mathscr{D}[v_0(n)] = \frac{VG_0G}{(q-1)\{1-KVG_0G\mathscr{D}[\mathrm{f}(n)]\}} = \theta_2(q) \quad . \quad (14)$$

Using eqn. (12) and substituting for case (3), we have

$$v_0(n) = VG_0G$$
$$\left[1 + K\sum_{r=0}^{r=n}\mathrm{f}(r)\sum_{\psi=0}^{\psi=n-1}v_0(n-r-\psi-1)\varepsilon^{-\alpha\psi} - \text{impulse at } t=0\right]$$

Taking discontinuous transforms of both sides,

$$\mathscr{D}[v_0(n)] = \frac{VG_0Gq}{q-1} + \frac{KVG_0G}{q}\mathscr{D}[\mathrm{f}(n)]\mathscr{D}\left[\sum_{\psi=0}^{\psi=n}v_0(n-\psi)\varepsilon^{-\alpha\psi}\right] - VG_0G$$

$$\mathscr{D}[v_0(n)] = VG_0G\left(\frac{q}{q-1}-1\right) + VG_0GK\mathscr{D}[\mathrm{f}(n)]\frac{\mathscr{D}[v_0(n)]}{q-\varepsilon^{-\alpha}}$$

$$\mathscr{D}[v_0(n)] = \frac{VG_0G(q-\varepsilon^{-\alpha})}{(q-1)\{q-\varepsilon^{-\alpha}-KVG_0G\mathscr{D}[\mathrm{f}(n)]\}} = \theta_3(q) \quad (15)$$

If the function of q, $\mathscr{F}(q) = \mathscr{D}[\mathrm{f}(t)]$, is substituted in $\theta_1(q)$, $\theta_2(q)$ or $\theta_3(q)$, the stability can be checked by observing whether the roots of the denominator lie within or outside the unit circle (inside corresponding to stability). Alternatively the inversion formula

$$v_0(n) = \frac{1}{2\pi i}\int_c \theta(q)q^{n-1}dq$$

may be used to give $v_0(n)$ as a function of discrete values of time. A plot similar to the Nyquist can be made using the denominator of $\theta(q)$.

(5) EXAMPLES OF RANGE MEASURING AND A.G.C. SYSTEMS

(5.1) Clamped-Gate System for Range Measuring (see Fig. 5)

Let

Delay-constant be D microsec/volt.

Gate-constant be H volts/microsec.

$T_1 = R_1C_1$ sec.

$T_2 = R_2C_2$ sec.

Then $DH = G$ (the gain of whole loop)

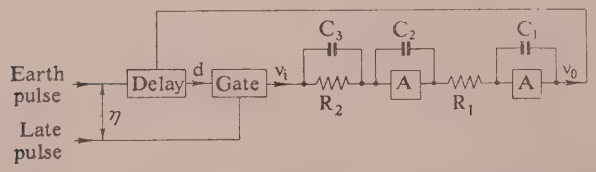


Fig. 5.—Block diagram of clamped-gate range-measuring unit.

The transfer function is seen to be given by

$$\mathrm{F}(p) = \frac{G}{pT_1}\left(\frac{1}{pT_2}+K\right) \text{ where } K = \frac{C_3}{C_2}$$

and since the input is a pulse, broadened to a step before entering the spoiled integrator, the function of p which we must consider from now is

$$\mathrm{F}(p) = G\frac{1}{p}\frac{1}{pT_1}\left(\frac{1}{pT_2}+K\right)$$

i.e. $$\mathrm{F}(p) = \frac{G}{T_1T_2p^3} + \frac{GK}{T_1p^2}$$

Hence $$\mathrm{f}(t) = \frac{Gt^2}{2!T_1T_2} + \frac{GK}{T_1}t$$

The discontinuous transform of $\mathrm{f}(t)$ is given by

$$\mathscr{D}[\mathrm{f}(n)] = \sum_{n=0}^{\infty}q^{-n}\mathrm{f}(n) = \sum_{n=0}^{\infty}q^{-n}\left(\frac{G}{2T_1T_2}n^2 - \frac{GKn}{T_1}\right)$$

and since $n^2 = n^{(2)} + n^{(1)}$ [using notation $n^{(m)} = n(n-1)\ldots(n-m+1)$]

we have

$$\mathscr{D}[\mathrm{f}(n)] = \frac{G}{2T_1T_2}\left[\sum_{n=0}^{\infty}q^{-n}n^{(2)} + \sum_{n=0}^{\infty}q^{-n}n^{(1)}\right] + \frac{GK}{T_1}\sum_{n=0}^{\infty}q^{-n}n^{(1)}$$

which, using tables of transforms, is equal to

$$\frac{G}{2T_1T_2}\left[\frac{2!q}{(q-1)^3} + \frac{q}{(q-1)^2}\right] + \frac{GK}{T_1}\frac{q}{(q-1)^2}$$

and this simplifies to

$$\mathscr{D}[\mathrm{f}(n)] = \frac{Gq}{(q-1)^2}\left[\frac{K}{T_1} + \frac{(q+1)}{2T_1T_2(q-1)}\right]$$

For the clamped gate, there is no discontinuous shaping $\phi(q)$, i.e. $\phi(q) = 1$, and substituting for $\mathscr{D}[\mathrm{f}(n)]$, $\phi(q)$ in eqn. (7),

$$\mathscr{D}[v_0(n)] = \mathscr{D}[v_i(n)]\frac{(q-1)\frac{Gq}{(q-1)^2}\left[\frac{K}{T_1}+\frac{(q+1)}{2T_1T_2(q-1)}\right]}{q+(q-1)\frac{Gq}{(q-1)^2}\left[\frac{K}{T_1}+\frac{(q+1)}{2T_1T_2(q-1)}\right]} \quad . \quad . \quad . \quad . \quad (7a)$$

Then eqn. (8) becomes

$$v_0(n) = \frac{1}{2\pi i}\int_c q^{n-1}\mathscr{D}[v_i(n)]\theta(q)dq \quad . \quad . \quad . \quad (8a)$$

c is contour around large circle centre at origin.

Now considering $v_i(n)$ to be in impulse at $t = 0$,

$$\mathscr{D}[v_i(n)] = 1$$

Eqn. (8a) simplifies to

$$v_0(n) = \frac{1}{2\pi i}\int_c q^{n-1}\frac{\left[\left(\frac{GK}{T_1}+\frac{G}{2T_1T_2}\right)q + \frac{G}{2T_1T_2} - \frac{GK}{T_1}\right]dq}{q^2+\left(\frac{GK}{T_1}+\frac{G}{2T_1T_2}-2\right)q+\left(\frac{G}{2T_1T_2}-\frac{GK}{T_1}+1\right)}$$

$$= \frac{1}{2\pi i}\int_c q^{n-1}\frac{N(q)}{M(q)}dq$$

Now $N(q)$ is of lower degree than $M(q)$.

Let us now consider

$$M(q) \equiv q^2 + \left(\frac{GK}{T_1}+\frac{G}{2T_1T_2}-2\right)q + \left(\frac{G}{2T_1T_2}-\frac{GK}{T_1}+1\right)$$

For stability, the roots of this must be less than unity in modulus, i.e.

$$\left|\frac{-\left(\frac{GK}{T_1}+\frac{G}{2T_1T_2}-2\right)\pm\sqrt{\left[\left(\frac{GK}{T_1}+\frac{G}{2T_1T_2}-2\right)^2-4\left(\frac{G}{2T_1T_2}-\frac{GK}{T_1}+1\right)\right]}}{2}\right|\leqslant 1$$

For critical damping, the roots must be equal,

i.e. $$\left(\frac{GK}{T_1}+\frac{G}{2T_1T_2}-2\right)^2=4\left(\frac{G}{2T_1T_2}-\frac{GK}{T_1}+1\right)$$

Equal roots when $$K=\sqrt{\left(\frac{4T_1}{GT_2}\right)}-\frac{1}{2T_2}$$

and they are less than unity in modulus if

$$K\leqslant\frac{4T_1}{G}-\frac{1}{2T_2}$$

Real roots when $$K>\sqrt{\left(\frac{4T_1}{GT_2}\right)}-\frac{1}{2T_2}$$

and they are less than unity in modulus if

$$K<\frac{2T_1}{G}-\frac{1}{2T_2}$$

Imaginary roots when

$$K<\sqrt{\left(\frac{4T_1}{GT_2}\right)}-\frac{1}{2T_2}$$

and they are less than unity in modulus if

$$K\geqslant\frac{1}{2T_2}$$

(5.2) **Summing-Gate System for Range Measuring** (see Fig. 6)

Its transfer function F(p) is the same as for the clamped-gate system,

i.e. $$\mathrm{F}(p)=\frac{G}{p}\frac{1}{pT_1}\left(\frac{1}{pT_2}+K\right)\quad K=\frac{C_3}{C_2}$$

Therefore the discontinuous transform of f(t) is the same,

i.e. $$\mathscr{D}[\mathrm{f}(n)]=\frac{Gq}{(q-1)^2}\left[\frac{K}{T_1}+\frac{(q+1)}{2T_1T_2(q-1)}\right]$$

But, for the summing-gate system, the shaping $\phi(q)$ is $q/(q-1)$. Hence eqn. (7) becomes, on substituting for $\theta(q)$ from eqn. (10),

$$\mathscr{D}[v_0(n)]=\frac{\mathscr{D}[v_i(n)]\,\mathscr{D}[\mathrm{f}(n)]}{1+\mathscr{D}[\mathrm{f}(n)]}$$

and $$v_0(n)=\frac{1}{2\pi i}\int_c q^{n-1}\frac{\mathscr{D}[v_i(n)]\,\mathscr{D}[\mathrm{f}(n)]}{1+\mathscr{D}[\mathrm{f}(n)]}dq$$

and if input is impulse at $t=0$, $D[v_i(n)]=1$.
Therefore

$$v_0(n)=\frac{1}{2\pi i}\int_c q^{n-1}\frac{\frac{Gq}{(q-1)^2}\left[\frac{K}{T_1}+\frac{(q+1)}{2T_1T_2(q-1)}\right]}{1+\left\{\frac{Gq}{(q-1)^2}\left[\frac{K}{T_1}+\frac{(q+1)}{2T_1T_2\,(q-1)}\right]\right\}}dq$$

$$=\frac{1}{2\pi i}\int_c q^{n-1}\frac{Gq\left[\frac{K}{T}(q-1)+\frac{1}{2T_1T_2}(q+1)\right]dq}{q^3+\left(\frac{GK}{T_1}+\frac{G}{2T_1T_2}-3\right)q^2+\left(\frac{G}{2T_1T_2}+3-\frac{GK}{T_1}\right)q-1}$$

This is stable if the denominator has roots less than unity in modulus.

If the roots are α, β, δ we see $\alpha\beta\delta=1$ and therefore, either $\alpha=\beta=\delta=1$ or at least one of them is greater than unity in modulus if the others are less than unity.

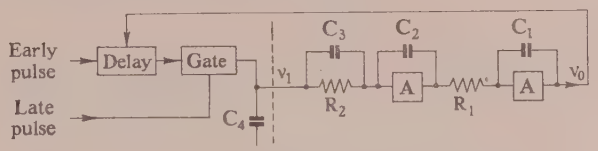


Fig. 6.—Block diagram of summing-gate range-measuring unit.

The case of $\alpha=\beta=\delta=1$ does not hold, however, since then

$$\frac{GK}{T_1}+\frac{G}{2T_1T_2}-3=-3$$

and $$\frac{G}{2T_1T_2}-\frac{GK}{T_1}+3=3$$

indicating that $G=0$, which is known to be untrue.
This system is therefore always unstable.

(5.3) **A.G.C. Systems**

The problem is to find the transfer function F(p) which gives the fastest response.

Consider the a.g.c. system operating as in arrangement (i) in Section 4.3. Let us take a simple *RC* consideration as the transfer function F(p) (Fig. 7).

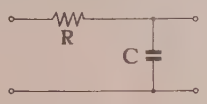


Fig. 7.—Simple *RC* filter as a.g.c. passive shaping device.

The transfer function of this system can be shown to be $1/pT+1$ (where $T=RC$), and since there is a gain K_1 (actually it is negative) in the feedback loop, we include this with the transfer function, so that

$$\mathrm{F}(p)=\frac{K_1}{pT+1}$$

whence

$$\mathrm{f}(t)=\frac{K_1}{T}\varepsilon^{-t/T}$$

and the discontinuous transform of this is given by

$$\mathscr{D}[\mathrm{f}(n)] = \sum_{n=0}^{\infty} q^{-n} \frac{K_1}{T} \varepsilon^{-n/T}$$

$$\mathscr{D}[\mathrm{f}(n)] = \frac{K_1}{T} \frac{q}{q - \varepsilon^{-1/T}}$$

Hence, substituting for $\mathscr{D}[\mathrm{f}(n)]$ in eqn. (13),

$$\mathscr{D}[v_0(n)] = \frac{VG_0Gq}{(q-1)\left(q - VG_0GK\frac{K_1}{T}\frac{q}{q-\varepsilon^{-1/T}}\right)}$$

and therefore

$$v_0(n) = \frac{1}{2\pi i}\int_c q^{n-1} \frac{VG_0Gq}{q(q-1) - (q-1)VG_0G\frac{K}{T}K_1\frac{q}{(q-\varepsilon^{-1/T})}}\,dq$$

For our a.g.c. system $VG_0G = 70$ volts, and $KK_1 = -\frac{1}{2}$.

Hence

$$v_0(n) = \frac{1}{2\pi i}\int_c q^{n-1} \frac{70(q-\varepsilon^{-1/T})}{(q-1)\left[(q-\varepsilon^{-1/T}) + \frac{35}{T}\right]}\,dq$$

and for stability the denominator of the integrand must have its roots equal to or less than unity in modulus;

i.e. $$\left|\varepsilon^{-1/T} - \frac{35}{T}\right| \leqslant 1$$

or $$\varepsilon^{-1/T} + 1 \geqslant \frac{35}{T} \geqslant \varepsilon^{-1/T} - 1$$

and $T \geqslant 20$ satisfies this.

Our particular pulse repetition frequency is 50 pulses/sec, i.e. 1/50 sec between pulses, and since, for preceding theory the time has been normalized, the actual time-constant of the network is equal to or greater than 20/50 sec for stability.

Let us now consider the response of this circuit to a negative impulse

$$\mathscr{D}[v_0(n)] = \frac{VG_0Gq}{(q-1)\{q - VG_0GK\mathscr{D}[\mathrm{f}(n)]\}}$$

$$= \frac{70(q-\varepsilon^{-1/T})}{(q-1)\left(q - \varepsilon^{-1/T} - \frac{35}{T}\right)}$$

Putting this in partial fractions, we have

$$v_0(n) = \frac{1}{2\pi i}\int_c q^{n-1}\left[\frac{A}{(q-1)} + \frac{B}{q - \left(\varepsilon^{-1/T} - \frac{35}{T}\right)}\right]dq$$

which evaluates to

$$A + B\left(\varepsilon^{-1/T} - \frac{35}{T}\right)^{n-1}$$

where $$A \simeq 2$$

and $$B \simeq 69$$

Hence $$v_0(n) = 2 + 69(-0{\cdot}8)^{n-1}$$

A diagrammatical representation of this output is given in Fig. 8, from which we see that an impulse of amplitude −70 volts decreases to 10% of its initial value in 12/50 sec, which is a reasonably rapid decay.

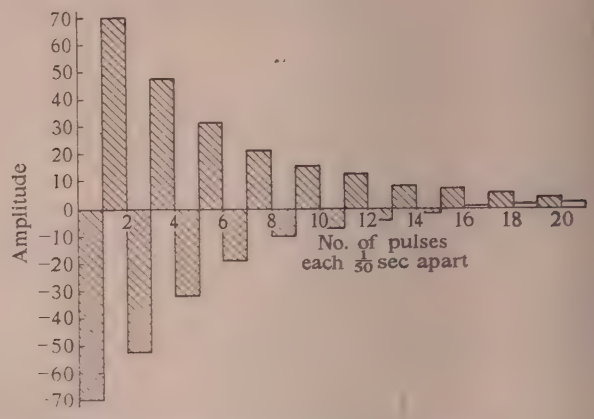


Fig. 8.—Diagrammatical representation of the output from gain-controlled i.f. amplifier with simple *RC* passive shaping.

Circuits such as that given in Fig. 9, with transfer functions given by

$$\mathrm{F}(p) = K_1 \frac{1 + T_3p}{p^2(T_1T_2 + T_1T_3) + p(R_1C_2 + T_1 + T_2 + T_3) + 1}$$

have been considered, and the one with the best response occurs when $R_3 = 0$.

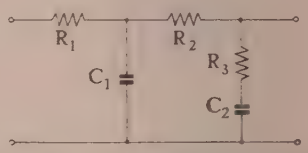


Fig. 9.—Alternative filter used for passive shaping in the a.g.c. loop.

The discontinuous transform of the output response to an impulse of amplitude −70 volts is given by

$$\mathscr{D}[v_0(n)] = \frac{VG_0Gq}{(q-1)\{q - VG_0GK\mathscr{D}[\mathrm{f}(n)]\}} \quad \cdots \quad (13)$$

$$= \frac{70(q-\varepsilon^a)(q-\varepsilon^b)}{(q-1)\mathrm{f}(q)}$$

where

$$a, b = \frac{-(R_1C_2 + T_1 + T_2) \pm \sqrt{[(R_1C_2 + T_1 + T_2)^2 - 4T_1T_2]}}{2T_1T_2}$$

$$T_1 = R_1C_1\,\mathrm{sec}$$

$$T_2 = R_2C_2\,\mathrm{sec}$$

and $$\mathrm{f}(q) = q^2 - q(\varepsilon^a + \varepsilon^b) + \frac{35(\varepsilon^b - \varepsilon^a)}{b-a} + \varepsilon^{a+b}$$

Then, taking $T_1 = T_2 = R_1C_2 = 5\,\mathrm{sec}$,

we have $$\mathscr{D}[v_0(n)] = \frac{70(q-0{\cdot}925)(q-0{\cdot}57)}{(q-1)(q-X)(q-Y)}$$

where X, Y are roots of $\mathrm{f}(q)$.

$$= \frac{A}{(q-1)} + \frac{B}{(q-X)} + \frac{C}{(q-Y)}$$

On evaluation,

$$A = \frac{70 - 105{\cdot}7 + 38{\cdot}22}{(1-X)(1-Y)}$$

$$B = \frac{70X^2 - X(105{\cdot}7) + 38{\cdot}22}{(X-1)(X-Y)}$$

and $$C = \frac{70Y^2 - Y(105\cdot7) + 38\cdot22}{(Y-1)(Y-X)}$$

so that

$$v_0(n) = \frac{1}{2\pi i}\int_c q^{n-1}\left[\frac{A}{(q-1)} + \frac{B}{(q-X)} + \frac{C}{(q-Y)}\right]dq$$

$$= A + BX^{n-1} + CY^{n-1}$$

$$= A + \left[\frac{(lX^2+mX+g)X^{n-1}}{(X-1)(X-Y)} + \frac{(lY^2+mY+g)Y^{n-1}}{(Y-1)(Y-X)}\right]$$

where $l = 70$

$m = -105\cdot7$

and $g = 38\cdot22$

$$= A + \left[\frac{-(Y-1)(lX^2+mX+g)X^{n-1} + (X-1)(lY^2+mY+g)Y^{n-1}}{(X-1)(Y-1)(Y-X)}\right]$$

$$= A + \left[\frac{-XYl(X^n-Y^n) - XYm(X^{n-1}-Y^{n-1}) - XYg(X^{n-2}-Y^{n-2}) + l(X^{n-1}-Y^{n-1}) + m(X^n-Y^n) + g(X^{n-1}-Y^{n-1})}{(X-1)(Y-1)(Y-X)}\right]$$

Now $(X^n - Y^n) = (X-Y)\sum_{r+s=n-1} X^rY^s$

and when X, Y are conjugate complex roots

$\sum_{r-s=n-1} X^rY^s$ is *always* real (n positive).

$$v_o(n) = A + \left[\frac{XYl\sum_{r+s=n-1} X^rY^s + Ym\sum_{r+s=n-2} X^rY^s + XYg\sum_{r+s=n-3} X^rY^s - l\sum_{r+s=n-2} X^rY^s - m\sum_{r+s=n-1} X^rY^s + g\sum_{r+s=n-2} X^rY^s}{XY-(X+Y)}\right]$$

for $n \geqslant 2$

But when $n = 1$,

$$v_o(n) = A + \frac{XYl + g - l(X+Y) - m}{XY-(X+Y)}$$

A diagrammatical representation of this output is given in Fig. 10.

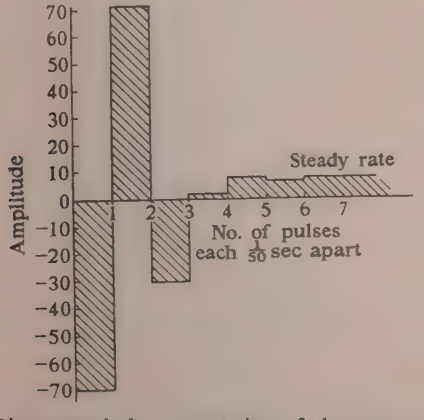


Fig. 10.—Diagrammatical representation of the output from gain-controlled i.f. amplifier with filter as shown in Fig. 9 with $R_3 = 0$.

Other networks, including "twin Ts," were tried, and we conclude that the best choice of network is that shown in Fig. 9, when $R_3 = 0$ with the appropriate C and R values. The time response to an impulse is seen to be faster than that for the simple RC circuit; in fact, the decay is complete in the time that 5 pulses have occurred, i.e. 0·1 sec.

The network in Fig. 9 was recommended[1] for use in a.g.c. systems operating with case (ii). When recommended it had not been tested, and the authors, for the sake of completeness, considered it. It was noted that as $T_3 = R_3C_3$ increased, the network became more stable, which agreed with the remarks in Reference 1.

(6) ACKNOWLEDGMENTS

The authors wish to thank Vickers-Armstrongs, Ltd., Mr. H. H. Gardner, Chief Designer, for permission to publish this work.

They are indebted to Mrs. K. A. Sellers and her staff for their valuable assistance in the computation of the results.

(7) REFERENCES

(1) Van Voorhis, N. S.: "Microwave Receivers" (McGraw-Hill Book Co.), M.I.T., Vol. 23, p. 407.

(2) Wiener, N.: "Generalized Harmonic Analysis," *Acta Mathematica*, 1930, **55**, p. 203.

(3) "Automatic and Manual Control" (Butterworth, 1952), p. 377.

(4) Cooper, S. R.: "A Study of a Second-order Sampling Servo," *Electronic Engineering*, 1953, **25**, p. 342.

(5) McColl, Leroy A.: "Fundamental Theory of Servo Mechanisms" (D. Van Nostrand, 1945).

(6) Milne, W. E.: "Numerical Calculus" (Princeton University Press, 1949).

(7) James, H. M., Nichols, N. B., and Phillips, R. S.: "Theory of Servo Mechanisms" (McGraw-Hill Book Co., 1947), M.I.T., Vol. 25.

(8) McLachlan, N. W.: "Complex Variable and Operational Calculus" (Cambridge University Press, 1947).

(9) Phillips, E. G.: "Functions of a Complex Variable" (Oliver and Boyd Mathematical Texts, 1940).

(10) Titchmarsh, E. C.: "Theory of Functions" (Oxford University Press, 1932).

(8) APPENDICES

(8.1) The Spectral Density of the Output from a Sampling Servo-Mechanism

In this Appendix an expression is obtained for the spectral density of the output from a sampling servo-mechanism. The sampling servo is of the general type described earlier; the expression for the output voltage from the servo is taken from eqn. (9) or (9*a*). The result obtained bears similarity to the equivalent result for the continuous case.

Analysis of the system including noise effect requires a knowledge of the output spectral density from a sampling servo-system in terms of some system operator and expressions for the input noise. The spectral density of the output of the system is obtained simply by taking the real part of the Fourier transform of the auto-correlation function of the output. The auto-correlation function of the output is obtained from the expression for the output of the system. The expressions in the following derivation are fairly lengthy, but the final equation is in a simple form and bears similarity with that for the continuous case.

Before considering the spectral density of the output of the sampling servo, the spectral density of sampled noise will be considered, to demonstrate the method.

Let the noise be represented by N(t) as shown in Fig. 11.

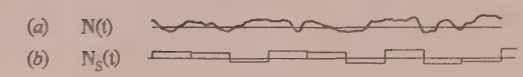


Fig. 11.—Waveforms.
(*a*) Noise waveform.
(*b*) Sampled noise waveform.

The sampling times are taken at unit times apart. The sampled noise is shown as $N_s(t)$; we wish to find the spectral density of the sampled waveform $N_s(t)$ in terms of some operation on the noise, $N(t)$.

The spectral density of $N_s(t)$ is given by the real part of the Fourier transforms of the auto-correlation function of $N_s(t)$, i.e.

$$S_s(\omega) = \mathscr{R}\int_{-\infty}^{+\infty} \varepsilon^{-i\omega t}\frac{1}{2T}\int_{-T}^{+T} N_s(r)N_s(t+r)\,dr\,dt$$
$$T \to \infty$$

It is convenient to make the substitutions: $t = m + \delta_1 t$ and $r = n + \delta_2 t$, where $\delta_1 t$ and $\delta_2 t$ are less than unity.
This gives

$$S_s(\omega) = \mathscr{R}\int_{-\infty}^{+\infty} \varepsilon^{-i\omega(m+\delta_1 t)}\frac{1}{2T}\int_{-T}^{T} N_s(n+\delta_2 t)N_s(m+\delta_1 t+n+\delta_2 t)\,d(n+\delta_2 t)\,d(m+\delta_1 t)$$
$$T \to \infty$$

$$S_s(\omega) = \mathscr{R}\int_{-\infty}^{+\infty} \varepsilon^{-i\omega(m+\delta_1 t)}\frac{1}{2T}\sum_{-T}^{+T}\int_{\delta_2 t=0}^{\delta_2 t=1} N_s(n+\delta_2 t)N_s(m+n+\delta_1 t+\delta_2 t)\,d\delta_2 t\,d(m+\delta_1 t)$$
$$T \to \infty$$

$$S_s(\omega) = \mathscr{R}\int_{-\infty}^{+\infty} \varepsilon^{-i\omega(m+\delta_1 t)}\left[\frac{1}{2T}\sum_{-T}^{+T}\int_{\delta_2 t=0}^{\delta_2 t=1-\delta_1 t} N_s(n+\delta_2 t)N_s(m+n+\delta_1 t+\delta_2 t)\,d\delta_2 t\right.$$
$$T \to \infty$$

$$\left. + \frac{1}{2T}\sum_{-T}^{+T}\int_{\delta_3 t=0}^{\delta_3 t=\delta_1 t} N_s(n+\overline{1+\delta_3 t-\delta_1 t})N_s(m+n+1+\delta_3 t)\,d\delta_3 t\right] d(m+\delta_1 t)$$
$$T \to \infty$$

In the second term $\delta_3 t = \delta_1 t + \delta_2 t - 1$

The integration with respect to $\delta_2 t$ and $\delta_3 t$ can now be carried out, giving

$$S_s(\omega) = \mathscr{R}\int_{-\infty}^{+\infty} \varepsilon^{-i\omega(m+\delta_1 t)}\{\psi(m)[1-\delta_1 t] + \psi(m+1)\delta_1 t\}\,d(m+\delta_1 t)$$

where
$$\psi(m) = \frac{1}{2T}\sum_{-T}^{+T} N(n)N(n+m)$$
$$T \to \infty$$

$$S_s(\omega) = \mathscr{R}\int_0^{\delta_1 t=1} \varepsilon^{-i\omega\delta_1 t}\left[\sum_{m=-\infty}^{m=\infty} \varepsilon^{-i\omega m}\psi(m)(1-\delta_1 t) + \sum_{m=-\infty}^{m=\infty} \varepsilon^{-i\omega m}\psi(m+1)\delta_1 t\right] d\delta_1 t$$

Substitution of $\sum_{m=-\infty}^{m=\infty} \varepsilon^{-i\omega m}\psi(m) = S_1(\omega)$ and $(m+1) = m_1$ gives

$$S_s(\omega) = \mathscr{R}\int_0^{\delta_1 t-1} \varepsilon^{-i\omega\delta_1 t}(1-\delta_1 t+\varepsilon^{i\omega}\delta_1 t)\,d\delta_1 t S_1(\omega)$$

$$S_s(\omega) = S_1(\omega)\frac{2}{\omega^2}(1-\cos\omega)$$

If the noise is white then $S_1(\omega) = 1$ and the spectral density for the sampled noise is given by

$$S_s(\omega) = \frac{2}{\omega^2}(1-\cos\omega)$$

This is interesting because it is the same as the spectral density of a random square-wave of amplitudes $+1$ and -1 only; i.e. if a coin is tossed and heads are plotted as $+1$, tails as -1, the spectral density of the waveform obtained is the same as that of the sampled white noise.[2]

We now proceed to find the spectral density of the output from a sampling servo-mechanism.

(8.2) Derivation of the Spectral Density of $N_0(t)$

The system under consideration is shown in Fig. 12.

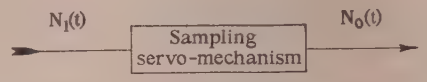


Fig. 12.—System under consideration.

The output from a sampling servo-mechanism in terms of the input is given by eqn. (9):

$$N_0(n+\delta t) = \frac{1}{2\pi i}\int_c \theta(q)\frac{\sum_r^r B_r\varepsilon^{\alpha_r\delta t}/q-\varepsilon^{\alpha_r}}{\sum_r^r B_r/q-\varepsilon^{\alpha_r}}\mathscr{D}[N_I(n)]q^{n-1}dq \quad . \quad . \quad . \quad . \quad (16)$$

where the contour c is a large circle around the origin, and

$$\mathscr{D}[N_I(n)] = \sum_{n=0}^{n=\infty} q^{-n}N_I(n)$$

$\theta(q)$, B_r and α_r are parameters describing the kind of units used within the sampling servo. $\theta(q)$ is a rational polynomial. B_r and α_r are constants; n is an integer and δt a time interval where $\delta t < 1$.

From eqn. (16), $N_0(n+\delta t)$ may be written as a convolution sum:

$$N_0(n+\delta t) = \sum_{s=0}^{s=n} g(s+\delta t)N_I(n-s) \quad . \quad . \quad (17)$$

where

$$g(n+\delta t) = \frac{1}{2\pi i}\int_c \theta(q)\frac{\sum_r^r B_r\varepsilon^{\alpha_r\delta t}/q-\varepsilon^{\alpha_r}}{\sum_r^r B_r/q-\varepsilon^{\alpha_r}}q^{n-1}dq \quad . \quad (18)$$

In eqn. (17) the lower limit $s = 0$ may be extended to minus infinity since

$$g(n+\delta_1 t) = 0 \text{ for } n < 0.$$

Also, if the noise is supposed to have started indefinitely far in the past, the upper limit must be extended to plus infinity; extending the limits eqn. (17) becomes

$$N_0(n+\delta t) = \sum_{s=-\infty}^{s=\infty} g(s+\delta t)N_I(n-s) \quad . \quad . \quad (19)$$

The general expression for the auto-correlation function of a time function is given by

$$\psi(t) = \frac{1}{2T}\int_{-T}^{+T} N(r)N(t+r)\,dr \quad . \quad . \quad . \quad (20)$$
$$\text{Lim } T \to \infty$$

Since eqn. (19) contains two variables instead of the single continuous variable, t, it is necessary to modify the form of eqn. (20) into the sum and an integral with respect to δt. Writing

he auto-correlation function for the output as $\psi_0(n + \delta_1 t)$, where t is integral and $\delta_1 t \leqslant 1$, we have

$$\psi_0(n + \delta_1 t) = \frac{1}{2T} \sum_{r=-T}^{r=T} \int_{\delta_2 t=0}^{\delta_2 t=1} N_0(r + \delta_2 t) N_0(r + n + \delta_1 t + \delta_2 t) d\delta_2 t \quad . \quad . \quad . \quad . \quad (21)$$

$$T \to \infty$$

In eqn. (21), $(\delta_1 t + \delta_2 t)$ may be greater than unity, and our expression for the output $N_0(n + \delta t)$ is for the total $\delta t \leqslant 1$; in order to overcome this difficulty the integral is split into two parts:

$$\psi_0(n + \delta_1 t) = \frac{1}{2T} \sum_{r=-T}^{r=T} \left[\int_{\delta_2=0}^{\delta_2 t=1-\delta_1 t} N_0(n + r + \overline{\delta_1 t + \delta_2 t}) N_0(r + \delta_2 t) d\delta_2 t \right.$$

$$\left. + \int_{\delta_3 t=0}^{\delta_3 t=\delta_1 t} N_0(n + r + 1 + \delta_3 t) N_0(r + \overline{\delta_3 t - \delta_1 t + 1}) d\delta_3 t \right] \quad . \quad (22)$$

If equations like eqn. (19) are substituted in eqns. (22), and eqn. (18) is substituted in eqn. (19), the real part of the Fourier transform of eqn. (22) gives the output spectral density. $\psi_0(n + \delta_1 t)$ is again in terms of two variables; thus the integral in $S_0(\omega)$ must be replaced by a sum and an integral with respect to $\delta_1 t$;

i.e.
$$S_0(\omega) = \sum_{n=-\infty}^{n=\infty} \varepsilon^{-i\omega n} \int_{\delta_1 t=0}^{\delta_1 t=1} \psi_0(n + \delta_1 t) \varepsilon^{-i\omega \delta_1 t} d\delta_1 t \quad . \quad (23)$$

Carrying out the substitution of eqn. (19) in eqn. (22) and substituting eqn. (22) in eqn. (23) gives

$$S_0(\omega) = \sum_{n=-\infty}^{n=\infty} \varepsilon^{-i\omega n}$$

$$\left\{ \int_{\delta_1 t=0}^{\delta_1 t=1} \frac{1}{2T} \sum_{r=-T}^{r=T} \left[\int_{\delta_2 t=0}^{\delta_2 t=1-\delta_1 t} \sum_{s_1=-\infty}^{s_1=\infty} N_1(n + r - s_1) g(s_1 + \overline{\delta_1 t + \delta_2 t}) \right.\right.$$

$$\times \sum_{s_2=-\infty}^{s_2=\infty} N_1(r - s_2) g(s_2 + \delta_2 t) d\delta_2 t$$

$$+ \int_{\delta_3 t=0}^{\delta_3 t=\delta_1 t} \sum_{s_1=-\infty}^{s_1=\infty} N_1(n + r + 1 - s_1) g(s_1 + \delta_3 t)$$

$$\left.\left. \times \sum_{s_2=-\infty}^{s_2=\infty} N_1(r - s_2) g(s_2 + \overline{\delta_3 t - \delta_1 t + 1}) d\delta_3 t \right] \varepsilon^{-i\omega \delta_1 t} d\delta_1 t \right\} . \quad (24)$$

In eqn. (24) the integration with respect to $\delta_1 t$, $\delta_2 t$ and $\delta_3 t$ can be carried out since g is known from eqn. (18). Assuming that the integration is performed and that $\int\int g g d\delta t_2 d\delta t_1$, etc., $= H$, eqn. (24) can be simplified to

$$S_0(\omega) = \sum_{n=-\infty}^{n=\infty} \varepsilon^{-i\omega n} \frac{1}{2T} \sum_{r=-T}^{r=T}$$

$$T \to \infty$$

$$\left[\sum_{s_1=-\infty}^{s_1=\infty} N_1(n + r - s_1) \sum_{s_2=-\infty}^{s_2=\infty} N_1(r - s_2) H_1(s_1, s_2) \right.$$

$$\left. + \sum_{s_1=-\infty}^{s_1=\infty} N_1(n + r + 1 - s_1) \sum_{s_2=-\infty}^{s_2=\infty} N_1(r - s_2) H_2(s_1, s_2) \right] . \quad (25)$$

where

$$H_1(s_1, s_2) = \int_{\delta_1 t=0}^{\delta_1 t=1} \varepsilon^{-i\omega \delta_1 t} \int_{\delta_2 t=0}^{\delta_2 t=1-\delta_1 t} g(s_1 + \overline{\delta_1 t + \delta_2 t}) g(s_2 + \delta_2 t) d\delta_2 t d\delta_1 t \quad . \quad (26)$$

and

$$H_2(s_1, s_2) = \int_{\delta_1 t=0}^{\delta_1 t=1} \varepsilon^{-i\omega \delta_1 t} \int_{\delta_3 t=0}^{\delta_3 t=\delta_1 t} g(s_1 + \delta_3 t) g(s_2 + \overline{\delta_3 t - \delta_1 t + 1}) d\delta_3 t d\delta_1 t$$

. . . . (27)

Writing
$$\psi_1(n) = \frac{1}{2T} \sum_{r=-T}^{r=T} N_1(r) N_1(r + n) \quad . \quad . \quad . \quad (28)$$

$$T \to \infty$$

eqn. (25) becomes

$$S_0(\omega) = \sum_{n=-\infty}^{n=\infty} \varepsilon^{-i\omega n} \sum_{s_1=-\infty}^{s_1=\infty} \sum_{s_2=-\infty}^{s_2=\infty} \psi_1(n + s_2 - s_1) H_1(s_1, s_2)$$

$$+ \psi_1(n + 1 + s_2 - s_1) H_2(s_1, s_2) \quad . \quad (29)$$

Substitution of
$$n + s_2 - s_1 = u$$

and
$$u + 1 = v$$

in eqn. (29) gives

$$S_0(\omega) = \sum_{u=-\infty}^{u=\infty} \varepsilon^{-i\omega u} \psi_1(u) \sum_{s_1=-\infty}^{s_1=\infty} \varepsilon^{-i\omega s_1} \sum_{s_2=-\infty}^{s_2=\infty} \varepsilon^{i\omega s_2}$$

$$[H_1(s_1, s_2) + \varepsilon^{i\omega} H_2(s_1, s_2)] \quad . \quad (30)$$

Writing
$$S_1(\omega) = \sum_{u=-\infty}^{u=\infty} \varepsilon^{-i\omega u} \psi_1(u) \quad . \quad . \quad . \quad . \quad (31)$$

we have

$S_0(\omega) = S_1(\omega)$ times a function of ω which depends only upon the servo mechanism.

In $H_2(s_1, s_2)$ substituion of $\delta_4 t = \delta_3 t - \delta_1 t + 1$ gives

$$H_2(s_1, s_2) = \int_{\delta_1 t=0}^{\delta_1 t=1} \varepsilon^{-i\omega \delta_1 t} \int_{\delta_4 t=1-\delta_1 t}^{\delta_4 t=1} g(\delta_1 t + \delta_4 t - 1 + s_1) g(s_2 + \delta_4 t) d\delta_4 t d\delta_1 t$$

. . . . (32)

Substitution for $H_1(s_1, s_2)$ and $H_2(s_1, s_2)$ in eqn. (30) and changing the order of integration and summation, we have

$$S_0(\omega) = S_1(\omega) \int_{\delta_1 t=0}^{\delta_1 t=1} \varepsilon^{-i\omega \delta_1 t} \int_{\delta_2 t=0}^{\delta_2 t=1} G(\omega, \delta_1 t + \delta_2 t) G(-\omega, \delta_2 t) d\delta_2 t d\delta_1 t$$

. . . . (33)

where
$$G(\omega, \delta t) = \sum_{s=-\infty}^{s=\infty} \varepsilon^{-i\omega s} g(s + \delta t)$$

The similarity between the sampling case and the continuous case is clearly seen in eqn. (33). The function $S_1(\omega)$ takes the place of the input spectral density, and the integrated product of $G(\omega, \delta_1 t + \delta_2 t)$ and $G(-\omega, \delta_2 t)$ corresponds to the normal $|Z(\omega)|^2$; eqn. (33) is not in a suitable form for evaluation since $(\delta_1 t + \delta_2 t)$ may be greater than unity. Eqn. (30) is in a suitable form where $\psi_1(u)$ is given by eqn. (31) and H_1 and H_2 are given by eqns. (26) and (27) together with eqn. (18).

The spectral density of the output from a sampling servo-mechanism is formed by the product of a periodic function with a period given by

$$\text{Period of } S_1(\omega) = \frac{2\pi}{\text{sampling time}}$$

and the function

$$\int_{\delta_1 t=0}^{\delta_1 t=1} \varepsilon^{-i\omega \delta_1 t} \int_{\delta_2 t=0}^{\delta_2 t=1} G(\omega, \delta_1 t + \delta_2 t) G(-\omega, \delta_2 t) d\delta_2 t d\delta_1 t$$

which is dependent upon the type of discontinuous and passive circuits used in the sampling servo-mechanism.

(8.3) Multiple Poles

If in the sampling servo-mechanism $F(p)$ contains multiple poles, the expression for $g(n + \delta t)$ can be obtained by taking $B_1 = B_2 = B_3$, etc., in eqn. (18), and $\alpha_1 \to \alpha_2 \to \alpha_3$, etc.

$g(n + \delta t)$ can be obtained more directly, however, by using the equation derived from eqn. (9*a*), when we have

$$g(n + \delta t) = \frac{1}{2\pi i}\int_{\rho} \frac{(q-1)\phi(q)\mathscr{D}[f(n+\delta t)]}{q + \phi(q)(q-1)\mathscr{D}[f(n)]} q^{n-1} dq$$

where $f(n)$ and $\phi(q)$ are as defined in the paper.

(8.4) Evaluation of Eqn. (30)

In evaluating eqn. (30) it was only necessary to take the summations with respect to s_1 and s_2 from zero to plus infinity. Great care must be exercised in obtaining the contribution from H_1 and H_2 when either s_1 or s_2 is zero.

(which is periodic) and the function of ω plotted in Fig. 13; this function is for a normalized pulse repetition frequency, and it is noted that the contribution to the noise-power level is zero at the repetition frequency and has a maximum (although very small) just before $\omega = 3\pi$. There will be smaller ones near all other multiples of π. The presence of these agrees directly with the example on spectral density for sampled noise given at the end of Section 8.1. There, it is seen that for "white" noise the spectral density of the output is

$$S_o(\omega) = \frac{2}{\omega^2}(1 - \cos\omega)$$

and maxima of this appear at points immediately before 3π, 5π, etc.

In the example considered, the time-constants employed were

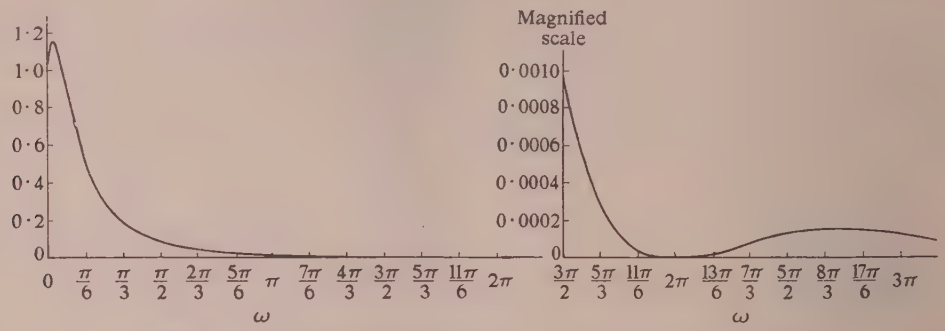


Fig. 13.—Plot of actual values calculated for range-measuring unit.

A practical range-measuring equipment was analysed from the stochastic standpoint; noise was considered to be applied to the input terminals of the unit (Fig. 5) and the problem was to determine the spectral density of the output noise in terms of the spectral density of the input noise.

Considerable work was necessary in order to apply the given formula (eqn. 30), and only the result and conclusions are given here.

(8.5) RESULTS

The output spectral density $S_o(\omega)$ is given by the product of

$$S_i(\omega) = \sum_{u=-\infty}^{\infty} \varepsilon^{-i\omega u}\psi(u)$$

unfortunately too large for the effect to be clearly illustrated; thus, because of their large smoothing effect, computations to an accuracy of 1 part in 10^6 had to be made before these maxima were detected.

If the input noise is "white"

$$S_i(\omega) = \sum_{u=-\infty}^{\infty} \varepsilon^{-i\omega u}\psi(u)$$

becomes a constant, and the function of ω in Fig. 13 then represents the output spectral density, $S_o(\omega)$, of the range-measuring unit.

621.317.335.3:621.372.413:621.3.015.3

Monograph No. 122 M
Mar. 1955

RESONANT-CAVITY MEASUREMENTS OF THE RELATIVE PERMITTIVITY OF A D.C. DISCHARGE

By K. W. H. FOULDS, Ph.D., B.Sc.(Eng.), Graduate.

(*The paper was first received 5th August, and in revised form 2nd December*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in March*, 1955.)

SUMMARY

The relative permittivity of a gaseous discharge has been measured at 2 100, 2 500, and 3 000 Mc/s, using an H_{011} resonant-cavity method. The preliminary experiments were carried out using a discharge in a mixture of argon ($2\frac{1}{2}$ mm Hg pressure) and mercury vapour. Later experiments employed a specially constructed low-pressure mercury-vapour tube.

The relative permittivity of each discharge decreases from unity almost linearly for increases in the discharge current. In the case of the low-pressure mercury-vapour discharge it has been observed that the Q-factor falls to a minimum at a particular discharge current which depends upon the resonant frequency of the cavity. It is suggested that this is due to an electron resonance phenomenon.

The detailed theory of the method of measurements is given in an appendix.

LIST OF PRINCIPAL SYMBOLS

c = Velocity of electromagnetic waves in free space.
e, m = Charge and mass of an electron, respectively.
E = Electric field strength.
$\hat{E}$ = Peak electric field strength.
$E_\phi(r)$ = Radial variation of electric field strength.
f = Impressed frequency.
δf = Small change in impressed frequency.
f_0 = Resonant frequency of the cavity.
H_z = Component of magnetic field strength along the z-axis.
$H_z(r)$ = Radial variation of axial magnetic field strength.
$\hat{H}$ = Peak magnetic field strength.
I_a = Current through the discharge tube.
J_e = Current density through the condenser due to the velocity of the electrons.
J_T = Total current density through the condenser.
k = Separation number, $k^2 = \gamma^2 + \omega^2\mu\epsilon$.
N = Electron density.
n = Order of the Bessel functions J(kr) and N(kr).
P = Power dissipated in the discharge.
Q = Q-factor $= \dfrac{2\pi f \times \text{Energy stored}}{\text{Power loss}}$
Q_0 = Q-factor of the cavity with no discharge current.
Q_L = Q-factor of the cavity for some specified discharge current.
r, ϕ, z = Axes of a cylindrical co-ordinate system.
T = Temperature of water bath.
W = Work required to fill the discharge region with electrons.
W_0 = Energy stored in the cavity with no discharge current.
ΔW = Change of energy stored in the region of the discharge.
$\nabla^2_{r\phi z} = \dfrac{\partial^2}{\partial r^2} + \dfrac{1}{r}\dfrac{\partial}{\partial r} + \dfrac{1}{r^2}\dfrac{\partial^2}{\partial \phi^2}$
γ = Propagation coefficient.
ϵ = Permittivity.
ϵ_0 = Permittivity of free space.
ϵ'_1 = Relative permittivity of the discharge.
$\Delta\epsilon'_1$ = Decrease of the relative permittivity of the discharge from unity.
θ_0 = Galvanometer deflection at resonance.
θ'_L = Galvanometer deflection at resonance for some specified discharge current.
λ = Wavelength of the impressed electromagnetic wave.
μ = Permeability of free space.
ν = Collisional frequency of the electrons.
σ = Conductivity of the discharge.

(1) BASIC PRINCIPLES

(1.1) Introduction

It is now well known that the ionosphere, or any other region containing free electrons, can have an appreciable effect upon the velocity of propagation of an electromagnetic wave passing through the region. If the electron density is not uniform, the electromagnetic wave is refracted. The basic principles of ionic refraction were first discussed by Eccles[1] in 1912, but Larmor[2] later showed that the refraction was due mainly to the light free electrons and not to the heavy positive ions. A comprehensive review of the propagation of electromagnetic waves through the ionosphere was given by Mimno[3] in 1937.

The manner in which the electromagnetic waves are refracted implies that the relative permittivity of the ionosphere is less than unity. This decrease can be illustrated quite easily in the following manner: Fig. 1 shows a parallel-plate condenser in

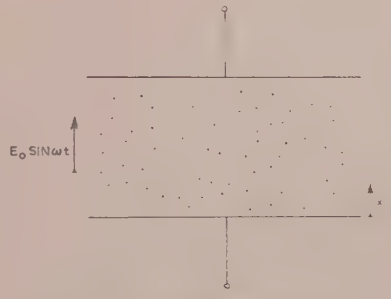


Fig. 1.—Parallel-plate condenser.

which the plates are separated by a low-pressure gas containing a uniform electron density of N electrons per unit volume. When an alternating electric field $E_0 \sin \omega t$ is applied between the plates, the electrons move under the action of the field. If the electrons collide with neutral particles, their radio-frequency energy is converted into random kinetic energy and power is abstracted from the electromagnetic field. Provided that the electron collisional frequency ν is large enough, the electrons

Correspondence on Monographs is invited for consideration with a view to publication.
Dr. Foulds is with Marconi's Wireless Telegraph Co., Ltd.
The paper is based on a thesis submitted for the degree of Doctor of Philosophy at London University.

experience a frictional force which can be represented by $mvdx/dt$.

The equation of motion for each electron is

$$m\frac{d^2x}{dt^2} + mv\frac{dx}{dt} = -eE_0 \sin \omega t$$

Solving this for dx/dt gives

$$\frac{dx}{dt} = \frac{-ev}{m(\omega^2 + \nu^2)}E + \frac{e}{m(\omega^2 + \nu^2)}\frac{dE}{dt} \text{ (where } E = E_0 \sin \omega t)$$

Each moving electron constitutes a current element, and the electron current density is given by

$$J_e = \frac{Ne^2\nu}{m(\omega^2 + \nu^2)}E - \frac{Ne^2}{m(\omega^2 + \nu^2)}\frac{dE}{dt} \quad . \quad . \quad (1)$$

To obtain the total current density J_t through the condenser, it is necessary to include the displacement current density $\epsilon_0(dE/dt)$ and eqn. (1) leads to

$$J_t = \frac{Ne^2\nu}{m(\omega^2 + \nu^2)}E + \left[\epsilon_0 - \frac{Ne^2}{m(\omega^2 + \nu^2)}\right]\frac{dE}{dt} \quad . \quad (2)$$

The first term in eqn. (2) represents the conduction current, and the second term represents the net displacement current which has to be supplied to the condenser plates from the external circuit. Writing σ and ϵ_1' for the conductivity and relative permittivity respectively of the region containing free electrons, this leads to

$$\sigma = \frac{Ne^2\nu}{m(\omega^2 + \nu^2)} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$\epsilon_1' = 1 - \frac{Ne^2}{m(\omega^2 + \nu^2)\epsilon_0} \quad . \quad . \quad . \quad . \quad . \quad (4)$$

If the collisional frequency ν is small compared with ω, the expression for ϵ_1' becomes

$$\epsilon_1' = 1 - \frac{Ne^2}{\omega^2 m\epsilon_0} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Eqns. (4) and (5) show that the relative permittivity of the ionized region decreases to less than unity by an amount depending upon the electron density and the frequency of the impressed electric field.

(1.2) Previous Experimental Investigations

Until recent years the permittivity of an ionized gas was generally measured by experiments in which a discharge formed the dielectric between the plates of a condenser. It was assumed that the change of capacitance of this condenser which was observed when the discharge current was varied was related in a simple manner to the change in the permittivity of the dielectric.

A large number of results obtained by this method has been reported in the technical literature (see Bibliography, items 4–12). Nearly all these experiments have shown that the relative permittivity of the discharge becomes less than unity only for very small currents. For larger discharge currents the relative permittivity reaches a well-defined minimum and thereafter increases steadily to a value greater than unity. Appleton and Childs[6] showed in a very convincing experiment that this anomalous behaviour was caused by the formation of a sheath of positive ions around each condenser plate.

Appleton and Chapman[7] repeated the experiments using redesigned apparatus which avoided the formation of positive ion sheaths. They then found that, provided that the discharge current was greater than a certain value, the capacitance of the test condenser and therefore the relative permittivity of the discharge, decreased approximately linearly with increase in discharge current. Assuming that the electron density increases linearly with discharge current, it is seen that this linear variation of ϵ_1' with I_a is indicated by eqn. (4).

Within recent years new techniques have been evolved for measuring the dielectric properties of materials in the microwave band of frequencies (see Bibliography, item 13). Adler[14] was the first to apply these microwave techniques to the measurement of the permittivity and the conductivity of a d.c. discharge. In these experiments a mercury-glow discharge tube was placed along the axis of a cylindrical cavity resonating in the E_{010} mode at a wavelength of 3 cm. The results show that the relative permittivity decreases linearly with discharge current, and the decrease is about 2% for a current of 0·4 mA.

(2) H_{011} RESONANT-CAVITY MEASUREMENTS UPON A MERCURY-VAPOUR AND ARGON DISCHARGE

(2.1) Description of the Cavity

Before the publication of Adler's paper it was decided to apply the resonant-cavity technique of measurements to a d.c. discharge.

The experiments were carried out on an 80-watt commercial lighting tube which was mounted along the axis of an H_{01} resonant cavity as shown in Fig. 2. An important characteristic

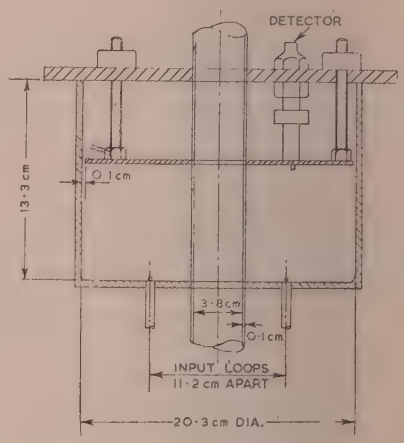


Fig. 2.—The resonant cavity.

of this mode is that the top of the cavity does not need to be in metallic contact with the walls. The open end of the cavity was therefore closed by a loosely fitting brass plunger, and the resonant frequency of the cavity was controlled by its position. A second characteristic of the H_{011} mode which makes it more convenient to use than the E_{010} mode used by Adler, is that there is negligible fringing of the electromagnetic field from the holes which accommodate the discharge tube.

The cavity was excited in the H_{011} mode by two small loops placed about half-way along diametrically opposite radii coupling with the radial magnetic field. The crystal detector unit, soldered on to the plunger about half-way along a radius, also coupled with the radial magnetic field.

(2.2) Investigation of the Mode Purity within the Cavity

The relation between the relative permittivity of the discharge and the shift in the resonant frequency of the cavity is established in Section 7.2, but this relation is valid only if the cavity resonates in an undisturbed H_{011} mode. If another resonant mode is also excited, the field pattern is no longer that of the undis-

turbed H_{011} mode and the relation no longer holds. Table 1 shows the unwanted modes which occur near the H_{011} resonant frequencies of 3 045, 2 500 and 2 117 Mc/s, at which it was proposed to carry out the experiments.

Table 1

RESONANT FREQUENCIES OF WANTED AND UNWANTED MODES

H_{011} resonant frequency	Unwanted mode and its resonant frequency							
Mc/s	Mc/s	Mc/s	Mc/s	Mc/s	Mc/s	Mc/s	Mc/s	Mc/s
3 045	E_{211} 3 448	E_{120} 3 297	H_{311} 3 141	E_{111} 3 045	$E_{\cdot\cdot0}$ 3 000	H_{211} 2 830	E_{011} 2 703	E_{020} 2 586
2 500	E_{310} 3 000	E_{211} 2 970	H_{311} 2 621	E_{020} 2 586	E_{111} 2 500	E_{210} 2 419	H_{211} 2 113	E_{011} 2 069
2 117	E_{020} 2 586	E_{210} 2 419	H_{311} 2 373	E_{111} 2 117	H_{211} 1 818	E_{110} 1 807		

The only modes likely to interfere are the H_{311}, E_{111} and E_{310} modes at approximately 3 045 Mc/s, the H_{311}, E_{020}, E_{111} and E_{210} modes at approximately 2 500 Mc/s, and the E_{111} mode at approximately 2 117 Mc/s.

The E_{111} and H_{011} modes always occur at the same resonant frequency in a perfect cylindrical cavity. The symmetrical input to the cavity, mentioned in the previous Section, was chosen as being the least likely to excite any of the unwanted modes. However, if any such modes were excited, they would be greatly damped by the air-gap between the plunger and the cavity walls.

The resonant frequencies given in Table 1 are those that would occur in an ideal cavity, but the values corresponding to the experimental cavity, including the discharge tube, may be substantially different. It is therefore necessary to make an experimental check on the purity of the field pattern within the cavity. When the plunger is rotated, the output loop couples with the radial magnetic field along each radius in turn. Any departure from the true H_{011} field-pattern is shown by a corresponding variation in the deflection of the associated galvanometer. The detector has a square-law characteristic; consequently a 10% variation in galvanometer deflection indicates a variation of about 5% in the radial magnetic field-strength. Before the experiments were begun, it was decided to tolerate up to $\pm 5\%$ variation of the magnetic field-strength from the mean.

Fig. 3 shows some typical graphs of galvanometer deflection against plunger rotation at each of the three resonant frequencies. The shapes cannot be easily explained, but the percentage variation of the deflection from the mean is within $\pm 10\%$.

(2.3) Experiments on a Low-Pressure Mercury-and-Argon Discharge

The frequency range 2 110–3 100 Mc/s is covered very conveniently by a cavity 20·3 cm in diameter and having a maximum inside height of 13·3 cm. Throughout the experiments, the resonant frequency of the cavity was measured with a wide-band wavemeter with a reading error within 1 part in 10^4. A CV67 klystron was used in the experiments at 3 000 Mc/s, and a CV273 triode, incorporated in a coaxial-line oscillator, for the experiments at the other frequencies.

The discharge was maintained by a 220-volt d.c. supply, and the current was controlled by suitable ballast resistors.

In some preliminary experiments using a smaller cavity

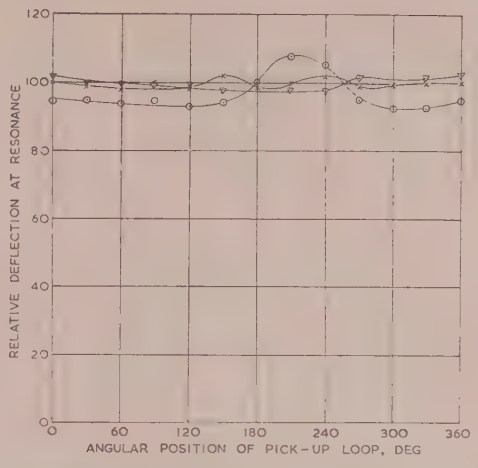


Fig. 3.—Variation of radial magnetic field.
○ $f_0 = 3\,036$ Mc/s.
× $f_0 = 2\,496$ Mc/s.
▽ $f_0 = 2\,114$ Mc/s.

(16·3 cm diameter), the resonant wavelength, the Q-factor and the deflection at resonance were each measured as a function of discharge current up to 300 mA. Fig. 4 shows some typical results for discharge currents up to 100 mA, and it is seen that,

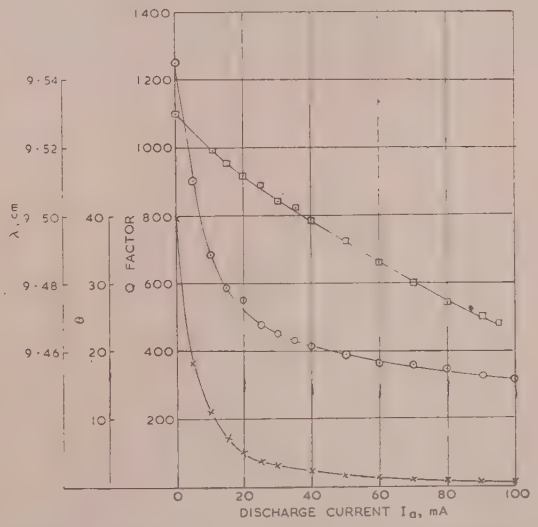


Fig. 4.—Variation of λ, Q and θ with discharge current.
⊡—⊡ Curve of λ, the resonant wavelength of the cavity.
○—○ Curve of Q, the Q-factor of the cavity.
×—× Curve of θ, the deflection at resonance.

whereas the resonant wavelength decreases linearly with discharge current, the Q-factor and the resonant deflection decrease very rapidly for small discharge currents, but thereafter decrease at a very much slower rate.

The majority of the experiments were carried out on the 20·3 cm-diameter cavity, and for discharge currents up to 0·8 amp. For these larger currents there was an appreciable change of temperature of the cavity, and it was necessary to perform each experiment as quickly as possible. Consequently only the resonant frequency of the cavity was measured for different discharge currents. It was observed, however, that the Q-

factor and resonant deflection continued to decrease slowly as indicated by Fig. 4.

Fig. 5 shows some typical results of the variations of the resonant frequency of the cavity as a function of discharge current. The apparent hysteresis effect is considered in Section 2.5.2.

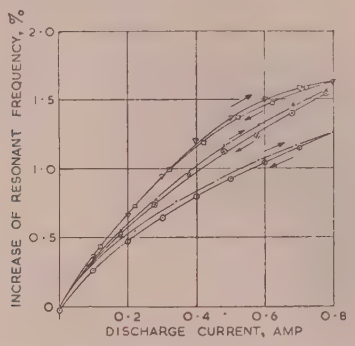


Fig. 5.—Variation of the resonant frequency of the cavity for increasing and decreasing currents.
△ and ⊡ $f_0 = 2\,114$ Mc/s.
× and ⊗ $f_0 = 2\,496$ Mc/s.
· and ⊙ $f_0 = 3\,036$ Mc/s.

(2.4) Results of the Resonant-Cavity Experiments

(2.4.1) Relative Permittivity.

Fig. 6 shows the relative permittivity of the discharge plotted as a function of current at the three frequencies. Since these experiments were of a preliminary nature, the relative permittivities have been calculated ignoring the effects of the glass wall [eqn. (29)]. These effects are not really negligible, and had they been taken into account the change in the relative permittivity would have been less than that shown in Fig. 6. The correction is a function of the frequency and amounts to about 35% at 3 000 Mc/s, to 28% at 2 500 Mc/s, and to 20% at 2 100 Mc/s. It is seen that the relative permittivity decreases linearly with the discharge current, and that for any given current it depends upon the impressed frequency; the lower the frequency the greater is the change in the relative permittivity from unity.

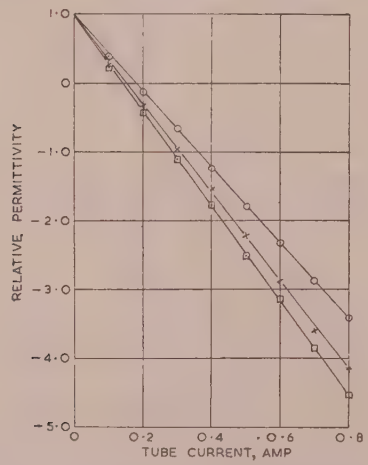


Fig. 6.—The relative permittivity of a discharge as a function of frequency and discharge current.
⊙ $f_0 = 3\,036$ Mc/s.
× $f_0 = 2\,496$ Mc/s.
⊡ $f_0 = 2\,114$ Mc/s.

The expression for the relative permittivity of an ionized gas has already been given in eqn. (4):

$$\epsilon_1' = 1 - \frac{Ne^2}{m(\omega^2 + \nu^2)\epsilon_0} \quad . \quad . \quad . \quad . \quad (4)$$

With the assumption already made that the electron density is proportional to the discharge current, the experimental results verify the above equation that the relative permittivity decreases linearly with increase in the electron density.

(2.4.2) Electron Collisional Frequency.

The electron collisional frequency, ν, can be calculated from the relative slopes of the curves of ϵ_1' versus I_a shown in Fig. 6.

Eqn. (4) can be rearranged to give

$$\omega^2 + \nu^2 = \frac{Ne^2}{m\epsilon_0}\frac{1}{\Delta\epsilon_1'} \quad . \quad . \quad . \quad . \quad . \quad (6)$$

It is seen that

$$\Delta\epsilon_1' = \frac{Ne^2}{m(\omega^2 + \nu^2)\epsilon_0}$$

which is the change in the relative permittivity from unity.

The values of $\Delta\epsilon_1'$ for a given discharge current can be read off the curves of Fig. 6, and therefore a family of straight lines can be drawn for ω^2 against $1/\Delta\epsilon_1'$ with the discharge current as parameter. These lines pass very nearly through a single point, giving an average value of $\nu = 2{\cdot}5 \times 10^{10}$ collisions/sec.

(2.4.3) Estimation of the Electron Density.

If the discharge is assumed to be uniform, eqn. (6) enables the electron density for any current to be calculated corresponding to the collisional frequency of $2{\cdot}5 \times 10^{10}$ collisions/sec. For a discharge current of 0·4 amp this gives $N = 7{\cdot}1 \times 10^{11}$ electrons/cm^3. These values of N and ν agree very satisfactorily with some results quoted by Denno,[15] measured from experiments based on the scattering of 3 cm wavelength electromagnetic waves by a discharge. The discharge was again produced by a commercial lighting tube, 3·25 cm in diameter, and for a discharge current of 0·4 amp the results quoted are as follows:

$$\nu = 4{\cdot}46 \times 10^{10} \text{ collisions/sec}$$
$$N = 7{\cdot}02 \times 10^{11} \text{ electrons/cm}^3$$

(2.5) Discussion of Phenomena Associated with this Technique of Measurement

During the course of the experiments reported above, two important phenomena associated with this resonant-cavity technique were discovered, and it is considered worth while discussing them here.

(2.5.1) Migration of Mercury Ions.

It is well known[16] that, when a unidirectional discharge current is passed through a mixture of argon and mercury vapour, the mercury ions gradually drift towards the cathode, and consequently near the anode the discharge takes place mainly through the argon. One of the assumptions made in this microwave method of measurement is that the electron density is proportional to the discharge current, and this will not be valid if there is a considerable migration of mercury ions towards the cathode. It was therefore decided to measure the resonant frequency of the cavity as a function of discharge current for a 50 c/s a.c. discharge.

The a.c. discharge causes two resonances and, as shown in Fig. 7(*a*), the frequency of one increases similarly to that for the d.c. discharge, but the frequency of the other increases only slightly with increase in current. When the current is less than

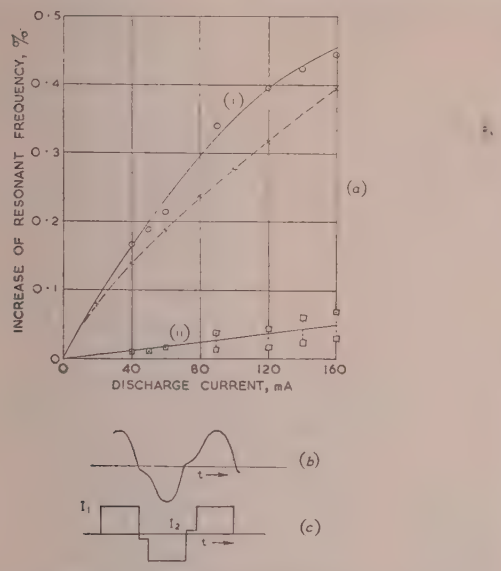


Fig. 7.—Cavity resonances and current waveforms for an a.c. discharge.
(*a*) Cavity resonances.
– – – D.C. discharge. ——— A.C. discharge.
(*b*) Typical current waveform of an a.c. discharge.
(*c*) Simplified current waveform of an a.c. discharge.

80 mA, the resonance curves are normal in shape, (ii) being considerably more damped than (i). For discharge currents greater than 80 mA the damped curves take the double-humped form associated with band-pass circuits. The frequency difference between the two humps is shown by the pairs of points joined by the dotted lines.

The explanation of the two resonances lies in the waveform of the discharge current, a typical waveform being shown in Fig. 7(*b*) (see Bibliography, item 17). For a rough approximation let it be assumed that the discharge current has the simplified stepped waveform of Fig. 7(*c*). During the part of the cycle for which the current is I_1 the cavity resonates at some frequency f_1, say, but when the current is I_2 the cavity resonates at a different frequency, f_2. The galvanometer cannot record these rapid fluctuations, and therefore when the impressed frequency is f_1 the galvanometer shows the average of the deflections corresponding to the discharge current I_1.

Similarly when the impressed frequency is f_2, the galvanometer shows the average of the deflections corresponding to the discharge current I_2. Thus for an increase in the discharge current f_1 increases at a rate similar to the d.c. discharge, and f_2 increases at a much slower rate. For the waveform shown in Fig. 7(*c*) the slopes of the lines f_1 versus I_a and f_2 versus I_a would be the ratio $I_1 : I_2$. This explanation fits the observed facts satisfactorily and indicates that multiple resonances will always occur if there are steps in the current waveform. One of the difficulties associated with the a.c. discharge is to determine the equivalent discharge current corresponding to an observed resonance. If the current could be made to have a square waveform, only one resonance would be obtained, and the corresponding discharge current would also be known. The relative permittivity for the discharge could be calculated from the measured change in the resonant frequency of the cavity by the same equations that are used for the d.c. discharge.

(2.5.2) Hysteresis Effect.

In all the experiments using discharge currents up to 0·8 amp, it was found that the resonant frequency of the cavity for a given discharge current depended upon whether the current was increasing or decreasing. This is shown very clearly in the curves of Fig. 5. The effect could be the result of a discharge phenomenon such as the formation of positive ion sheaths at the wall of the discharge tube, or a hysteresis effect between the electron density and discharge current, or it could be a thermal effect such as expansion of the cavity and glass tube. Both by calculation and by direct measurement it has been shown conclusively that the hysteresis effect is primarily, if not completely, due to the time lag between the temperature of the apparatus and the current in the discharge.

(3) RESONANT CAVITY MEASUREMENTS. LOW-PRESSURE MERCURY-VAPOUR DISCHARGE

The experiments on the commercial lighting-tube have shown that the collisional frequency of the electrons has an appreciable effect upon the permittivity of the discharge in the frequency range 2 000–3 000 Mc/s. The resonant-cavity experiments were therefore repeated upon a discharge at a pressure sufficiently low to ensure that the collisional frequency would be negligible compared with that of the impressed electromagnetic field. The most convenient type of discharge satisfying this requirement is one with a mercury-pool cathode and containing no gases apart from the mercury vapour. Discharge tubes of this kind have been described by Langmuir[18] and Killian[19] and by other workers making probe studies on discharges.

(3.1) Description of the Discharge Tube

Fig. 8 is a diagram of the low-pressure tube which was made of suitable low-loss glass. The discharge between the anode A and the mercury-pool cathode replaces the fluorescent tube in

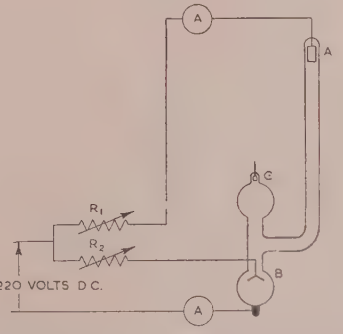


Fig. 8.—Circuit diagram of the low-pressure mercury-vapour tube.

the cavity. The discharge between the anode B and the cathode is an auxiliary "keep-alive" system, the anode being a flat cone to prevent blasts of mercury entering the long vertical arm. There is a third anode *C*, but this was not used in the experiments reported here. Below the level of C the discharge tube was immersed in a water bath whose temperature was controlled to within $\pm \frac{1}{4}$°C. To make certain that the mercury vapour did not condense on the wall of the long tube, this tube was surrounded by a hot-air jacket. The hot air also circulated through the cavity, maintaining it at an even temperature, and in this manner the thermal hysteresis described in Section 2.5.2 was avoided.

The currents to the anodes A and B were controlled independently, and during each test the total cathode current was kept constant, generally at 2 amp.

(3.2) Preliminary Experiments

(3.2.1) Asymmetrical Resonance Curves.

The initial experiments using this low-pressure discharge tube showed that for discharge currents greater than 50 mA the

resonance curves of the cavity were asymmetrical; a brief account of this phenomenon has been given by Cullen and Foulds.[20] The experiments have been extended, and Fig. 9 shows some typical curves for discharge currents up to 1·0amp for several water-bath temperatures.

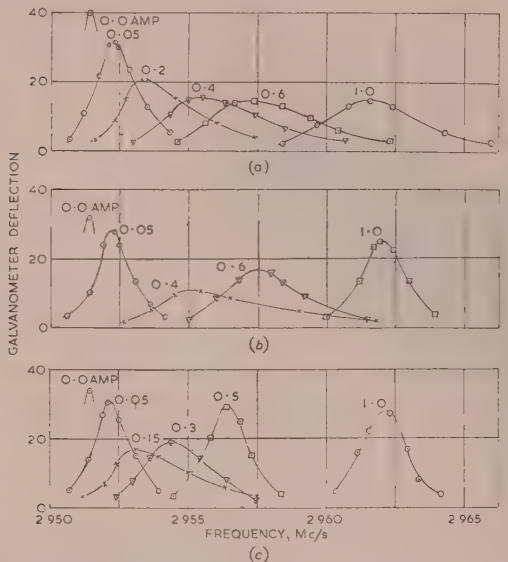


Fig. 9.—Resonance curves of the cavity as a function of discharge current and temperature of the water bath.
(*a*) $T = 20°C$. (*b*) $T = 29°C$. (*c*) $T = 36°C$.

It is seen that the higher the temperature of the water bath, i.e. the greater the pressure within the tube, the lower is the discharge current at which the asymmetry and damping occur. This suggests that the asymmetry and damping occur over a limited range of electron density.

In an attempt to obtain more information about the asymmetrical resonance curves, the cavity was made tunable over a small frequency band by means of a tuning ring. The resonance curves of the cavity could then be obtained at any specified frequency within the band by tuning the cavity, and the curves obtained in this manner were almost the same as those in Fig. 9.

(3.2.2) **Fluctuations in a D.C. Discharge.**

Fluctuations of the discharge current would result in the cavity being swept in and out of resonance, and the resonance curve indicated by the galvanometer would then appear damped. The output voltage from the detector was therefore displayed on a cathode-ray oscillograph, and this showed immediately that the crystal current was fluctuating. At the same time, the fluctuations of the light intensity from the discharge were examined by a photocell, and the output from this was also displayed on the oscillograph (a double-beam instrument). Fig. 10 shows some pairs of traces of crystal current and photocell output for a discharge current of 0·3amp at different points on the average resonance curve. By comparing the galvanometer deflections with the input frequency, it is seen that the average resonance curve is appreciably asymmetrical. The traces make it quite clear that the fluctuations in the crystal current on the high-frequency side of the curve are different from those on the low-frequency side. It is more correct to state that fluctuations similar to those on the low-frequency side of the curve do occur on the high-frequency side, but they are almost completely masked by the irregular fluctuations right up to resonance! At frequencies less than the apparent resonant frequency f_0, the

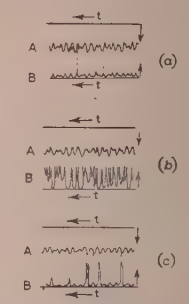


Fig. 10.—Correlation between average crystal current, instantaneous current, and instantaneous light intensity.
(*a*) Frequency $= f_0 - 3·6$Mc/s, average deflection = 10.
(*b*) Frequency $= f_0$(2 950Mc/s), average deflection = 53.
(*c*) Frequency $= f_0 + 6·5$Mc/s, average deflection = 10.
A = Photocell output, B = Instantaneous deflection.

fluctuations in the two traces correspond quite closely with the crystal-current fluctuations occurring either after a decrease in the light intensity or before the following increase. At frequencies greater than f_0, the traces show that the cavity is still frequently swept right up to, and indeed through, resonance. These crystal-current pulses occur irregularly and appear either just after an increase in the light intensity or before the next decrease. The traces for the high-frequency side of the resonance curve do not make it clear why some of the light-intensity fluctuations are associated with large crystal-current pulses and some are not.

The fluctuations in the discharge occur even though the discharge is supplied from an accumulator, and they clearly have a profound effect upon measurements of this kind.

An extensive experimental study has been made of these fluctuations by using two similar photocells to trace them as they travel through the discharge. The investigation has shown that the complicated nature of the light-intensity fluctuations is the result of the interaction of two sets of striae which move in opposite directions. Those which travel towards the cathode have been called the positive striae, and they are almost sinusoidal. The others, which have been called the negative striae, travel towards the anode and appear to occur at random. Table 2 shows the velocities of the striae for different water-bath temperatures.

Table 2

VELOCITY OF STRIAE

Temperature of water bath	Positive striae	Negative striae
°C	10^5 cm/sec	10^5 cm/sec
20	2	3·7
29	1·9	3·7
36	1·5	2

(3.2.3) **Summary.**

The asymmetry and damping of the resonance curves are certainly associated with the fluctuations in the light intensity of the discharge, but the complete explanation of the asymmetry has not been found from these experiments. It is probable that the crystal-current fluctuations are caused by corresponding fluctuations in the electron density. The bright striae are caused by large numbers of electrons being brought to rest by collision, and the momentary increases of the electron density above the mean may be very much larger than the, perhaps more regular, decreases below the mean. A detailed examination of the

fluctuations of the light intensity has shown the existence of two oppositely travelling sets of striae, and the increases in the electron density will be very marked at the points where the two sets of striae interact.

Although it was not known at the time of the experiments, this phenomenon of moving striae had already been extensively examined by Donahue and Dieke.[21] Their experiments also showed the existence of two sets of striae, travelling in opposite directions, but in their case both sets were periodic. Their paper puts forward a qualitative theory to account for the motion of the striae.

(3.3) Description and Results of the Resonant-Cavity Measurements

(3.3.1) Description of Experiments.

The resonant-cavity experiments using the low-pressure mercury-vapour discharge were carried out at approximately the same frequencies as those on the commercial discharge-tube, i.e. at 2 100, 2 500 and 3 000 Mc/s. For these experiments the cavity was excited by a single loop coupling with the longitudinal magnetic field tangential to the wall. Before each series of measurements at the different resonant frequencies, considerable care was taken to ensure that the cavity resonated in a substantially undisturbed H_{011} mode. At each frequency, when the plunger was rotated, the deflection varied less than 10% from the mean.

The experimental procedure was to measure the resonant frequency of the cavity and the height of the resonance curve, as a function of discharge current up to 2·0 amp. In order to measure the true height of the resonance curve, even for the range of current in which the fluctuations occurred, the curves were displayed on an oscillograph.

The experiments showed that the change in the resonant frequency of the cavity was almost a linear function of the discharge current, with no hysteresis effect.

(3.3.2) Relative Permittivity and Collisional Frequency.

It will be remembered that the primary reason for carrying out the experiments upon the low-pressure mercury-vapour discharge was that the collisional frequency of the electrons should then be negligible compared with the impressed angular frequency. With the usual assumption that at a given pressure the electron density is proportional to the discharge current, the graph of the relative permittivity and discharge current for a constant water-bath temperature should form a series of straight lines whose slopes are inversely proportional to ω^2. The experiments were carried out at two water-bath temperatures, and it was therefore expected that the experimental results would form two such sets of straight lines. Bearing this in mind, the experimental results of ϵ_1' versus I_a shown in Fig. 11 are surprising. It is seen immediately that, although the curves are approximately linear, the slopes do not conform to the $(1/\omega)^2$ relation, neither is there very much difference between the curves for the two temperatures. The relative permittivity of the argon and mercury-vapour discharge did not satisfy the $(1/\omega)^2$ relationship either, and this was explained by the relatively large electron collisional frequency ν. If the departure from the $(1/\omega)^2$ relation shown by the results for the low-pressure tube is also to be explained by electron collisions, the required value of ν is $1{\cdot}5 \times 10^{10}$ collisions/sec, irrespective of water-bath temperature. This does not seem to be a very likely explanation, because such a large value of ν is associated with a much larger amount of damping than was observed.

An alternative estimation of the electron collisional frequency can be based upon the results given by the dotted curves of Fig. 12, which shows the experimental relationship between the deflection at resonance and the discharge current. For convenience the

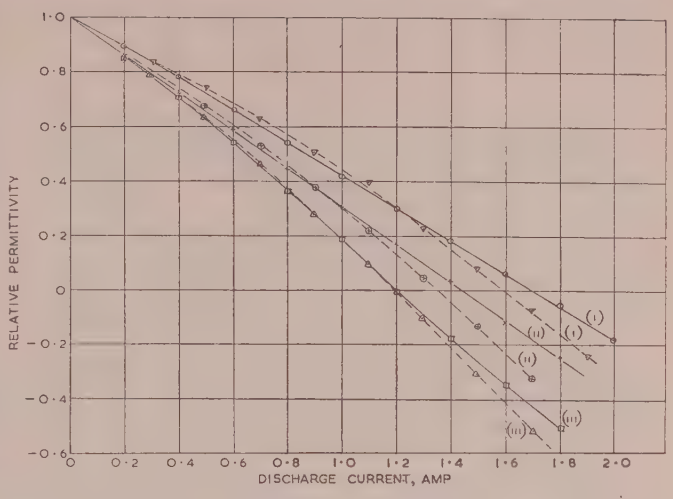


Fig. 11.—The relative permittivity of the discharge as a function of discharge current, water-bath temperature and frequency.
(i) $f_0 = 2\,946$ Mc/s, (ii) $f_0 = 2\,500$ Mc/s, (iii) $f_0 = 2\,119$ Mc/s.
——— $T = 20°$C. – – – – $T_0 = 38{\cdot}6°$C.

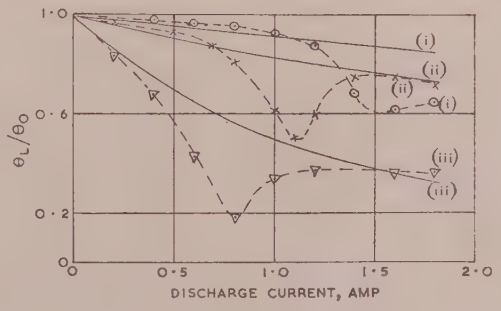


Fig. 12.—Curves of θ_L/θ_0 versus I_a.
$T = 38{\cdot}6°$C.
(i) $f_0 = 2\,946$ Mc/s, (ii) $f_0 = 2\,500$ Mc/s, (iii) $f_0 = 2\,119$ Mc/s.
——— Theoretical. – – – – Experimental.

deflections have been plotted as a fraction of the resonant deflection for no discharge current. Each curve shows a well-defined dip which occurs at a discharge current that depends upon the impressed frequency. Results were obtained both by altering the input frequency and by tuning the cavity for a fixed frequency. Although the two sets of curves are not identical, the main characteristics are very similar.

At each of the three impressed frequencies it was found that the current corresponding to the minimum resonant deflection depended upon the exact resonant frequency of the cavity. In the experiments at 2 100 Mc/s, for example, an increase in the resonant frequency of only 14 Mc/s caused the critical current to increase from 0·65 to 0·84 amp. The results for the water-bath temperature of 20°C are very little different from those shown in Fig. 12, the critical currents being generally a little greater.

If the current in the input loop to the cavity is constant, the deflection at resonance is proportional to the square of the Q-factor. The curves of Fig. 12 may suggest, therefore, that the Q-factor of the cavity decreases approximately uniformly when the discharge current is increased, but that there is superimposed upon this a pronounced dip caused by some unspecified phenomenon. Section 7.3.4 shows that the collisional frequency of

the electrons, ν, is related to the regular decrease in the Q-factor by the equation

$$\nu = \frac{\omega}{2}\frac{f}{\delta f}\frac{1}{Q_0}\left(\frac{Q_0}{Q_L} - 1\right) \quad . \quad . \quad . \quad . \quad (7)$$

For a discharge current of 1·8amp the values deduced for ν from the experimental data are $2{\cdot}88 \times 10^8$ collisions/sec at 3 000Mc/s, $1{\cdot}06 \times 10^8$ collisions/sec at 2 500Mc/s, and $0{\cdot}83 \times 10^8$ collisions/sec at 2 100Mc/s.

In the experiments at 3 000Mc/s the discharge current of 1·8amp is within the region of the pronounced minimum, and it is to be expected that the value of ν would be too great. At the other frequencies the current of 1·8amp is well removed from the minimum, and the two values agree satisfactorily.

It is very interesting to calculate the Q-factor of the cavity as a function of discharge current directly from the corresponding graphs of the change in the resonant frequency. Eqn. (7) can be rearranged to give

$$\frac{Q_L}{Q_0} = \frac{1}{1 + 2Q_0\frac{\nu}{\omega}\frac{\delta f}{f}} \quad . \quad . \quad . \quad . \quad . \quad (8)$$

This leads to the equation

$$\frac{\theta_L}{\theta_0} = \left(\frac{1}{1 + 2Q_0\frac{\nu}{\omega}\frac{\delta f}{f}}\right)^2 \quad . \quad . \quad . \quad . \quad (9)$$

The full lines of Fig. 12 show the curves of θ_L/θ_0 versus I_a calculated from this equation corresponding to $\nu = 0{\cdot}94 \times 10^8$ collisions/sec. It is seen that, apart from the region involving the minimum, the two sets of curves of Fig. 12 agree satisfactorily. This lends support to the suggestion that the initial decrease in the Q-factor is a direct result of the conductivity of the discharge, whereas the well-defined minimum is the result of some other phenomenon.

The electron collisional frequency is almost independent of the water-bath temperature. Furthermore the curves of ϵ_1' versus I_a (Fig. 11), are again almost independent of the temperature. This suggests that the mean free path (m.f. path) remains nearly constant over the temperature range 20°–38·6°C. At 20°C the m.f. path of electrons in mercury vapour is about 12cm, which is much greater than the tube radius. At 40°C the m.f. path is about 2·5cm, which is approximately equal to the radius of the discharge tube. Most of the absorption of the microwave power is contributed by those electrons which are in the strongest electric field when they lose their energy; this will be near the walls of the discharge tube. It seems possible that at the lower temperature the effective m.f. path of the electrons, so far as absorption is concerned, is limited by the physical size of the discharge tube; and on the average an electron would have a m.f. path of about the tube radius, i.e. about 2cm. Killian[19] showed that the mean random velocity of the electrons is about 10^8cm/sec, and if the m.f. path is limited to the tube radius the collisional frequency becomes $0{\cdot}5 \times 10^8$ collisions/sec, which is not very different from the value deduced above. However, Killian found evidence from his experiments that the m.f. path could be greater than the tube diameter (6·2cm) in spite of the collisions with the walls.

An alternative way of accounting for the similarity between the results for water-bath temperatures of 20°C and 38·6°C is to stipulate that the conditions within the discharge do, in fact, remain almost independent of the ambient temperature. This would be the case if the tube contained unwanted gas at a pressure considerably greater than the saturated vapour pressure of the mercury. It has already been pointed out that the corresponding damping of the resonance curves would have been very much greater than that observed. More definite proof that this is not the explanation of the results is that for a constant discharge current it was observed that the potential difference between the anode A and the cathode gradually increased as the water-bath temperature was increased. For example, for a current of 0·4amp the potential difference increased from 78 volts at 20°C to 97 volts at 38·6°C. This shows that the pressure within the discharge tube increases noticeably with the increase of water-bath temperature, and therefore the saturated vapour pressure of the mercury is not being swamped by the pressure of some unwanted gas.

(3.3.3) Estimation of the Electron Density.

Eqn. (6) enables the average electron density to be calculated for any discharge current, provided that the electron collisional frequency is known. One of the original purposes of this research was to compare the results deduced from these resonant-cavity measurements with those published by Killian.[19] Table 3 gives values of electron density for an average current density of 0·166amp/cm². They have been calculated for both $\nu = 1{\cdot}5 \times 10^{10}$ and $0{\cdot}94 \times 10^8$ collisions/sec.

Table 3

ELECTRON DENSITY

Source	Electron density	
	At a water-bath temperature of 20°C	At a water-bath temperature of 38·6°C
	electrons/cm³	electrons/cm³
Quoted by Killian	$0{\cdot}72 \times 10^{11}$	$1{\cdot}51 \times 10^{11}$
Cavity measurements $\nu = 1{\cdot}5 \times 10^{10}$ collisions/sec	$1{\cdot}96 \times 10^{11}$	$2{\cdot}15 \times 10^{11}$ ($2{\cdot}88 \times 10^{11}$)
Cavity measurements $\nu = 0{\cdot}94 \times 10^8$ collisions/sec		
$f_0 = 2\,946$Mc/s	$1{\cdot}21 \times 10^{11}$	$1{\cdot}31 \times 10^{11}$ ($1{\cdot}76 \times 10^{11}$)
$f_0 = 2\,500$Mc/s	$1{\cdot}02 \times 10^{11}$	$1{\cdot}17 \times 10^{11}$ ($1{\cdot}57 \times 10^{11}$)
$f_0 = 2\,119$Mc/s	$0{\cdot}92 \times 10^{11}$	$0{\cdot}97 \times 10^{11}$ ($1{\cdot}30 \times 10^{11}$)

Of the two figures quoted in the results for 38·6°C, the top figure is derived directly from the measured change in the resonant frequency of the cavity assuming a uniform discharge. The lower figure, in brackets, is derived by applying a multiplying factor of 1·34 to allow for a parabolic distribution of the electron density. This distribution is discussed in Section 7.3.3.

It is not surprising that the results do not agree with those of Killian. The value $\nu = 1{\cdot}5 \times 10^{10}$ collisions/sec is necessary to make the electron density constant for a given current density when calculated at different frequencies, and it is about 150 times greater than the estimated value. The corresponding electron density is therefore not a reliable estimate of the electron density actually occurring in the discharge. The value $\nu = 0{\cdot}94 \times 10^8$ collisions/sec gives different electron densities for a given current density when calculated at different frequencies. This shows either that the equation for ϵ_1' is different from eqn. (4) or that eqn. (25) used to calculate the relative permittivity from the change in the resonant frequency of the cavity does not hold in these experiments with this discharge tube.

(3.4) Explanation of the Anomalous Deflection

There are several possible explanations of the pronounced minimum that occurs in the graph (Fig. 12) of resonant deflection against discharge current, and these are discussed below.

(3.4.1) Electron Resonance.

Bearing in mind that at each resonant frequency the maximum absorption occurred over a small range of discharge current, depending upon the actual experimental conditions, it is found that the discharge current varies approximately as the square of the applied frequency. This recalls the electron-oscillation phenomenon suggested by Tonks and Langmuir.[25] They show that electron oscillations can be expected when

$$\frac{Ne^2}{\omega^2 m\epsilon_0} = 1$$

This criterion corresponds to the relative permittivity of the region being zero, with the collisional frequency ν much smaller than ω. If for a given electron density the electrons have a natural tendency to oscillate at a specified frequency, such oscillations will become of an appreciable amplitude when the electrons are excited by an impressed electric field of the same frequency. When the r.f. energy of these electrons is converted into random kinetic energy by collisions, there is an appreciable increase in the absorption of power within the cavity.

Absorption arising in this manner is somewhat similar to that found by Appleton and Childs,[6] but in their experiments the critical frequency was the gyro-magnetic frequency, i.e. the frequency at which the electrons spin around the superimposed axial lines of magnetic flux.

An explanation based on the electron resonance can be made to fit the qualitative results of the experiments. If the discharge is not uniform over its cross-section, there is a range of current over which the critical electron density exists, and over this range the absorption within the discharge is very much greater than that corresponding to the normal conductivity term

$$\sigma = \frac{Ne^2\nu}{m(\omega^2 + \nu^2)}$$

If the discharge were completely uniform, the critical electron density would occur at only one value of discharge current and the absorption would be very severe.

No direct proof of this electron resonance has been obtained, but it seems a feasible explanation of the observed experimental results. The existence of such an electron resonance would also explain why the relative permittivities of the discharge measured at different frequencies do not follow the $(1/\omega)^2$ relation. If, within the discharge, there is an annulus in which the electrons oscillate vigorously under the r.f. electric field, the inner part of the discharge will be shielded from the electric field. Under these conditions it is not possible to calculate the relative permittivity of the discharge in the manner indicated in Section 7.2.

(3.4.2) Alternative Explanations.

(*a*) *Variation of Klystron Output.*

The changes in the deflection at resonance could have been caused by corresponding changes in the power output from the klystron. It was shown, however, by measuring the bandwidth of the resonance curves, that the Q-factor varied approximately as $\sqrt{\theta}$, indicating that the klystron output remained constant during the measurements.

(*b*) *Interference from Unwanted Modes.*

The experiments of Rommel[23] and the theoretical work of Kaiser and Closs[24] have shown that, if the electric vector of an electromagnetic wave is normal to the axis of a gaseous discharge, appreciable absorption occurs at certain electron densities. In the present experiments the electric field could be normal to the discharge only if the cavity were resonating in a mode different from the H_{011} one. This is considered to be very unlikely because great care was taken to ensure that no unwanted modes were present.

(*c*) *A Skin-Effect Phenomenon.*

A skin-effect phenomenon is to be expected in experiments such as these. The total loss occurring in the discharge is given by the integral

$$\int_{vol.\ of\ discharge} \sigma E^2 dv$$

where

$$\sigma = \frac{Ne^2\nu}{m(\omega^2 + \nu^2)}$$

For small discharge currents the conductivity is low, and there is little loss in the discharge. This loss increases for increases in discharge current. For large discharge currents the conductivity becomes high and the electric field hardly penetrates the discharge. The loss in the discharge region therefore begins to decrease again for further increase in the discharge current. However, calculations have shown that this skin effect does not play an important part in the explanation of the curves.

(4) CONCLUSION

It has been shown that the resonant frequency of an H_{011} resonant cavity is increased by the presence of the discharge and that it is possible to calculate the permittivity of the discharge required to cause this increase. Using the H_{011} resonant-cavity method it is therefore possible to measure the effective relative permittivity of the discharge under these conditions. However, the method is not in itself sufficient to measure the electron density in the discharge, because the change in the resonant frequency of the cavity is a function of both the total number of electrons and their distribution within the discharge. A further complication has been observed that, when the electron collisional frequency is small, there appears to be an electron resonance phenomenon which results in an anomalous absorption.

(5) ACKNOWLEDGMENTS

The paper has been based upon work carried out by the author in the Electrical Engineering Laboratories at University College, London. The author would like to thank Prof. H. M. Barlow for the active interest he has always shown in this work, and Dr. A. L. Cullen, who has made many helpful suggestions. The author would also like to thank the National Physical Laboratory for the loan of a wide-band wavemeter, and finally the Department of Scientific and Industrial Research for the provision of a maintenance grant.

(6) BIBLIOGRAPHY

(1) Eccles, W. H.: "On the Diurnal Variation of the Electric Waves occurring in Nature, and on the Propagation of Electric Waves round the Bend of the Earth," *Proceedings of the Royal Society*, 1912, **A87**, p. 79.

(2) Larmor, J.: "Why Wireless Electric Rays can bend round the Earth," *Philosophical Magazine*, 1924, **48**, p. 1025.

(3) Mimno, H. R.: "The Physics of the Ionosphere," *Review of Modern Physics*, 1937, **9**, p. 1.

(4) Van der Pol: Thesis, University of Utrecht, 1920.

(5) Gutton, H., and Clement, J.: "Dielectric Properties of Ionised Gases," *Comptes rendus hebdomadaires des séances de l'Academie des sciences*, 1927, **184**, p. 441, and "The Propagation of Electromagnetic Waves round the Earth," *ibid.*, 1927, **184**, p. 676.

(6) APPLETON, E. V., and CHILDS, E. C.: "On Some Radio-Frequency Properties of Ionized Air," *Philosophical Magazine*, 1930, **10**, p. 969.
(7) APPLETON, E. V., and CHAPMAN, F. W.: "The Collisional Frequency experienced by Vibrating Electrons in Ionized Air," *Proceedings of the Physical Society*, 1932, **44**, p. 246.
(8) BANNERJEE, S. S.: "On the Resonance Frequency of Oscillatory Circuits with Leaky Condenser and its Bearing on Measurement of the Dielectric Constant of Ionized Gas," *Philosophical Magazine*, 1934, **17**, p. 834.
(9) MITRA, S. K., and BANNERJEE, S. S.: "Dielectric Constant of Ionized Air," *Nature*, 1935, **136**, p. 512.
(10) IMAM, A., and KHASTGIR, S. R.: "Dielectric Constant of Ionized Gases," *Philosophical Magazine*, 1937, **23**, p. 858.
(11) GANGOPADHYAYA, S., and KHASTGIR, S. R.: "Dielectric Constant of Ionized Gases," *ibid.*, 1938, **25**, p. 883.
(12) KHASTGIR, S. R., and RAHMAN, S. M. F.: "The Dielectric Constant and Electrical Conductivity of Gases and Vapours Ionized by X-Rays at Ultra-High Radio Frequency," *ibid.*, 1940, **29**, p. 353.
(13) HORNER, F., TAYLOR, T. A., DUNSMUIR, R., LAMB, J., and WILLIS JACKSON: "Resonance Methods of Dielectric Measurement at Centimetre Wavelengths," *Journal I.E.E.*, 1946, **93**, Part III, p. 53.
(14) ADLER, F. P.: "Measurement of the Complex Conductivity of an Ionized Gas at Microwave Frequencies," *Journal of Applied Physics*, 1949, **20**, p. 1125.
(15) DENNO, S. N.: Ph.D. Thesis, University of Liverpool, 1951.
(16) COTTON, H.: "Electric Discharge Lamps" (Chapman and Hall, 1946).
(17) FRANCIS, V. J.: "Fundamentals of Discharge Tube Circuits" (Methuen and Co., Ltd., 1948).
(18) LANGMUIR, I.: "Positive Ion Currents in the Positive Column of the Mercury Arc," *General Electric Review*, 1923, **26**, p. 731.
(19) KILLIAN, T. J.: "The Uniform Positive Column of an Electric Discharge in Mercury Vapour," *Physical Review*, 1930, **35**, p. 1238.
(20) CULLEN, A. L., and FOULDS, K. W. H.: "Resonant Cavity Measurements on a D.C. Discharge," *Nature*, 1952, **169**, p. 236.
(21) DONAHUE, T., and DIEKE, G. H.: "Oscillatory Phenomena in Direct Current Glow Discharges," *Physical Review*, 1951, **81**, p. 248.
(22) DENNO, S. N., PRIME, H. A., and CRAGGS, J. D.: "The Scattering of 3cm Radiation by Ionized Gases," *Proceedings of the Physical Society*, 1950, **63B**, p. 726.
(23) ROMMEL, D.: "Radio Reflections from a Column of Ionized Gas," *Nature*, 1951, **167**, p. 243.
(24) KAISER, T. R., and CLOSS, R. L.: "Theory of Radio Reflections from Meteor Trails," *Philosophical Magazine*, 1952, **43**, p. 1.
(25) TONKS, L., and LANGMUIR, I.: "Oscillations in Ionized Gases," *Physical Review*, 1929, **33**, p. 195.
(26) PINCHERLE, L.: "Electromagnetic Waves in Metal Tubes filled longitudinally with Two Dielectrics," *ibid.*, 1944, **66**, p. 118.
(27) COLLIE, C. H., HASTED, J. B., and RITSON, D. M.: "The Cavity Resonator Method of Measuring the Dielectric Constant of Polar Liquids in the Centimetre Band," *Proceedings of the Physical Society*, 1948, **60**, p. 71.
(28) ROSEN, P.: "The Propagation of Electromagnetic Waves in a Tube containing a Coaxial D.C. Discharge," *Journal of Applied Physics*, 1949, **20**, p. 868.
(29) British Association Mathematical Tables, Volume 6 (Cambridge University Press, 1937).
(30) MACLEAN, W. R.: "The Resonator Action Theorem," *Quarterly of Applied Mathematics*, 1945, **2**, p. 329.
(31) BARLOW, H. E. M., and CULLEN, A. L.: "Microwave Measurements" (Constable and Company Ltd., 1950).

(7) APPENDIX

(7.1) The Detailed Theory of the Experimental Measurements

Several workers have shown that a knowledge of the resonant frequency of a cylindrical cavity containing a coaxial dielectric rod can be used to evaluate the permittivity of the dielectric (see Bibliography, items 13, 14, 26, 27 and 28). Among these, Pincherle[26] deduces the conditions which must be satisfied for an H_{01} wave to be propagated down a cylindrical guide which has an axial dielectric rod.

The problem which has to be solved when the "dielectric rod" is an electric discharge is more complicated because the discharge has to be contained in a glass tube. The analysis which is given in Section 7.2 extends the two-media results of Pincherle to include the effect of the glass tube, i.e. to the case of three media, still assuming that each medium is homogeneous and loss free.

An approximate condition for resonance corresponding to an inhomogeneous discharge is described in Section 7.3.3. The analysis is based on the principle of the adiabatic invariance of action, and ignores the glass wall. Finally Section 7.3.4 leads to an approximate relation between the Q-factor of the cavity and the collisional frequency of the electrons in the discharge.

(7.2) Condition for H_{011} Resonance in a Cavity containing Three Homogeneous and Loss-Free Dielectrics

(7.2.1) Rigorous Solution.

Fig. 13 shows a cylindrical cavity containing three concentric, homogeneous and loss-free dielectrics.

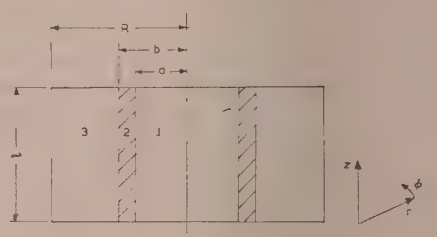


Fig. 13.—A cylindrical cavity.

The longitudinal component of the magnetic field H_z must satisfy the equation

$$\nabla^2_{r\phi z} H_z = -\mu\epsilon \frac{\partial^2 H_z}{\partial t^2} \quad . \quad . \quad . \quad . \quad (10)$$

Assume that

$$H_z = H_z \varepsilon^{-\gamma z} \varepsilon^{j\omega t}$$

where H_z is a function of r and ϕ only. Eqn. (10) must be satisfied in each of the three media, and therefore

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2}\right) H_z + (\gamma^2 + \omega^2 \mu_m \epsilon_m) H_z = 0$$

where μ_m and ϵ_m are the permeability and permittivity, respectively, of the medium being considered. The permeabilities are

approximately the same as that of free space, and the wave equation in each medium can therefore be expressed as

$$\nabla^2_{r\phi} H_z + k_m^2 H_z = 0 \quad . \quad . \quad . \quad . \quad (11)$$

where

$$\left.\begin{aligned} k_m^2 = k_1^2 &= \gamma^2 + \omega^2\mu\epsilon_1 \text{ for medium 1 (the discharge)} \\ &= k_2^2 = \gamma^2 + \omega^2\mu\epsilon_2 \text{ for medium 2 (the glass wall)} \\ &= k_3^2 = \gamma^2 + \omega^2\mu\epsilon \text{ for medium 3 (the air space)} \end{aligned}\right\} \quad (12)$$

To prevent confusion arising between the field components in the different media, it is essential to adopt some unambiguous notation. The field components in the different media are recognized by the number of primes attached to the letters E or H. Thus $E'_\phi(r)$, $E''_\phi(r)$, $E'''_\phi(r)$ are the electric fields at any radius r in media 1, 2 and 3, respectively.

Consider then eqn. (11) applied to any medium

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2}\right) H_z + k^2 H_z = 0 \quad . \quad . \quad (13)$$

This equation is most conveniently solved by assuming that H_z is the product of two independent functions $R(r)$, a function of r only, and $\Phi(\phi)$ a function of ϕ only. Eqn. (13) can then be separated into two ordinary differential equations, and the solution is

$$H_z = [AJ_n(kr) + BN_n(kr)](A_1\varepsilon^{jn\phi} + B_1\varepsilon^{-jn\phi}) \quad . \quad (14)$$

A, B, A_1, B_1 and n are constants determined by the boundary conditions. The field pattern within the cavity is circularly symmetrical and therefore $n = 0$. Consider from now onwards only the radial variation of the field components: then

$$\left.\begin{aligned} H_z(r) &= AJ_0(kr) + BN_0(kr) \\ H_r(r) &= \frac{\gamma}{k}[AJ_1(kr) + BN_1(kr)] \\ E_\phi(r) &= \frac{-j\omega\mu}{k}[AJ_1(k_1 r) + BN_1(kr)] \end{aligned}\right\} \quad . \quad . \quad (15)$$

The constants A and B can be expressed in terms of the electric field $E_\phi(L)$ and the longitudinal magnetic field $H_z(L)$, if these are known at a known radius $r = L$.

The equations for $H_z(r)$ and $E_\phi(r)$ then become

$$H_z(r) = \frac{1}{J_0(kL)N_1(kL) - J_1(kL)N_0(kL)} \left\{\begin{aligned} &H_z(L)[N_1(kL)J_0(kr) - J_1(kL)N_0(kr)] \\ &+ \frac{k}{j\omega\mu}E_\phi(L)[N_0(kL)J_0(kr) - J_0(kL)N_0(kr)] \end{aligned}\right\} \quad . \quad (16)$$

$$E_\phi(r) = \frac{1}{J_0(kL)N_1(kL) - J_1(kL)N_0(kL)} \left\{\begin{aligned} &E_\phi(L)[N_1(kr)J_0(kL) - J_1(kr)N_0(kL)] \\ &- \frac{j\omega\mu}{k}H_z(L)[N_1(kL)J_1(kr) - J_1(kL)N_1(kr)] \end{aligned}\right\} \quad . \quad (17)$$

Eqns. (15), (16) and (17) can be used to give the relative magnetic and electric field strengths in each of the three media of the cavity problem; this enables the condition for resonance to be obtained.

The H_z and E_ϕ field components in the air space are

$$H'''_z(r) = A_3J_0(k_3r) + B_3N_0(k_3r) \quad . \quad . \quad . \quad (18)$$

$$E'''_\phi(r) = \frac{-j\omega\mu}{k_3}[A_3J_1(k_3r) + B_3N_1(k_3r)] \quad . \quad (19)$$

E'''_ϕ is tangential to the metal wall and therefore must be zero at the wall. Putting $r = R$ in eqn. (19) and equating $E'''_\phi(R) = 0$, B_3 is obtained in terms of A_3. The expressions for H'''_z and E'''_ϕ at the boundary between the glass and air ($r = b$), are given by eqns. (18) and (19) by writing B_3 in terms of A_3, and putting $r = b$. The equations are

$$H'''_z(b) = A_3\left[J_0(k_3b) - \frac{J_1(k_3R)}{N_1(k_3R)}N_0(k_3b)\right] \quad . \quad . \quad (20)$$

$$E'''_\phi(b) = \frac{-j\omega\mu}{k_3}A_3\left[J_1(k_3b) - \frac{J_1(k_3R)}{N_1(k_3R)}N_1(k_3b)\right] \quad . \quad (21)$$

The electric field and the longitudinal magnetic field are each continuous at this boundary.

Therefore $E''_\phi(b) = E'''_\phi(b)$

and $H''_z(b) = H'''_z(b)$

In medium 2, E''_ϕ and H''_z are now known at a specified radius. It is therefore possible to calculate E''_ϕ and H''_z at any radius r in this medium, by substituting the known expressions for $E''_\phi(b)$ and $H''_z(b)$ in eqns. (16) and (17) for the quantities $E_\phi(L)$ and $H_z(L)$, respectively. The fields at the boundary between the discharge and glass are then obtained by substituting $r = a$. The expressions for the field at this boundary are

$$E''_\phi(a) = \frac{-j\omega\mu A_3}{J_0(k_2a)N_1(k_2a) - J_1(k_2a)N_0(k_2a)} \frac{1}{N_1(k_3R)}([1] + [2]) \quad . \quad . \quad . \quad . \quad (22)$$

$$H''_z(a) = \frac{A_3}{J_0(k_2a)N_1(k_2a) - J_1(k_2a)N_0(k_2a)} \frac{1}{N_1(k_3R)}([3] + [4]) \quad . \quad . \quad . \quad . \quad (23)$$

where

$$[1] = \frac{1}{k_3}[J_1(k_2a)N_0(k_2b) - N_1(k_2a)J_0(k_2b)][J_1(k_3R)N_1(k_3b) - J_1(k_3b)N_1(k_3R)]$$

$$[2] = \frac{1}{k_2}[N_1(k_2a)J_1(k_2b) - J_1(k_2a)N_1(k_2b)][N_0(k_3b)J_1(k_3R) - J_0(k_3b)N_1(k_3R)]$$

$$[3] = [N_0(k_2a)J_1(k_2b) - J_0(k_2a)N_1(k_2b)][N_0(k_3b)J_1(k_3R) - J_0(k_3b)N_1(k_3R)]$$

$$[4] = \frac{k_2}{k_3}[N_0(k_2a)J_0(k_2b) - J_0(k_2a)N_0(k_2b)][N_1(k_3b)J_1(k_3R) - J_1(k_3b)N_1(k_3R)]$$

The corresponding electric and magnetic fields in medium 1 at the boundary are

$$E'_\phi(a) = \frac{-j\omega\mu}{k_1}A_1J_1(k_1a)$$

$$H'_z(a) = A_1J_0(k_1a)$$

The fields are again continuous at the boundary, and therefore these expressions can be equated to eqns. (22) and (23). Eliminating the constants A_1 and A_3 leads to the equation which has to be satisfied at resonance,

i.e. $$\frac{J_1(k_1a)}{k_1J_0(k_1a)} = \frac{[1] + [2]}{[3] + [4]} \quad . \quad . \quad . \quad . \quad (24)$$

It will be convenient later to refer to this equation as

$$P(k_1a) = Q(k_3b) \quad . \quad . \quad . \quad . \quad . \quad (25)$$

(7.2.2) **Approximate Solution.**

The complicated expression $Q(k_3b)$ can be simplified by noting that the glass wall is thin. By writing $b = a + \delta r$ the Bessel functions can be expanded by Taylor's theorem, and the approximate expression for $Q(k_3b)$ becomes

$$S(k_3b) = \frac{T(k_3b) - \delta r}{\left(1 - \frac{\delta r}{a}\right) + k_2^2 \delta r T(k_3b)} \quad . \quad . \quad (26)$$

where $$T(k_3b) = \frac{1}{k_3} \frac{N_1(k_3b)J_1(k_3R) - J_1(k_3b)N_1(k_3R)}{N_0(k_3b)J_1(k_3R) - J_0(k_3b)N_1(k_3R)}$$

If the wall becomes vanishingly thin the problem reduces to the two-media problem studied by Pincherle.[26] In the limit when $\delta r \to 0$, eqn. (25) becomes

$$P(k_1a) = T(k_3a) \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

i.e. $$\frac{J_1(k_1a)}{k_1J_0(k_1a)} = \frac{1}{k_3} \frac{N_1(k_3a)J_1(k_3R) - J_1(k_3a)N_1(k_3R)}{N_0(k_3a)J_1(k_3R) - J_0(k_3a)N_1(k_3R)} \quad . \quad (29)$$

This is exactly the same relation that Pincherle deduces.

There is one important difference between the two relations $P(k_1a) = S(k_3b)$ and $P(k_1a) = T(k_3a)$, i.e. how each depends upon the frequency. If $T(k_3a)$ is plotted against k_3, the curve is independent of frequency. The graph of $S(k_3b)$ versus k_3, however, does depend upon the frequency because the denominator of $S(k_3b)$ contains the term in k_2^2, and for any value of k_3, k_2 depends upon the frequency. The physical explanation of this difference is that the electric field tends to concentrate into the glass and thus alters the electric field within the cavity. The higher the frequency the more does the electric field concentrate into the glass wall, and consequently the curve of $S(k_3b)$ versus k_3 is frequency dependent.

(7.2.3) **A Graphical Solution to the Characteristic Equation.**

The most convenient way of dealing with eqn. (24) is to plot $P(k_1a)$ versus k_1a and $Q(k_3b)$ versus k_3a and measure graphically the corresponding pairs of k_3a and k_1a which make $P(k_1a) = Q(k_3b)$. From these graphs it is convenient to draw the curve of $(k_3a)^2 - (k_1a)^2$ versus $(k_3a)^2$, and it is then necessary to relate the change in k_3a to the measured change in the resonant frequency of the cavity corresponding to a known discharge current. The measured change in f_0 is small, and it is sufficiently accurate to say

$$\delta(k_3a)^2 = \frac{\partial(k_3a)^2}{\partial\omega}\delta\omega$$

and from eqn. (12)

$$\delta(k_3a)^2 = 2\left(\frac{a}{c}\right)^2 \omega_0 \delta\omega \quad . \quad . \quad . \quad . \quad (30)$$

Eqn. (12) also leads to

$$k_3^2 \quad k_1^2 = \left(\frac{\omega}{c}\right)^2 (1 - \epsilon_1')$$

and this gives

$$\epsilon_1' = 1 - \left(\frac{c}{a\omega}\right)^2 [(k_3a)^2 - (k_1a)^2] \quad . \quad . \quad . \quad (31)$$

The calculation of the relative permittivity for any given current is now a straightforward matter. The measured change in the resonant frequency is substituted in eqn. (30) to give the value of $\delta(k_3a)^2$. The value of $(k_3a)^2 - (k_1a)^2$ is found from the graph, and this is substituted directly into eqn. (31) to give the relative permittivity. It was found most convenient to plot graphs of the relative permittivity of the discharge corresponding to any measured shift in the resonant frequency of the cavity. The points on these graphs were computed from the values of the Bessel functions tabulated in the published Tables.[29]

(7.3) Inhomogeneous Discharge

(7.3.1) **Introduction.**

The condition for resonance expressed in eqn. (24) is true only for a uniform dielectric. Killian,[19] however, has shown that the radial distribution of the electron density for an electric discharge at a pressure of 0·0055 mm Hg is approximately parabolic. An approximate condition is deduced below which must be satisfied for the cavity to resonate in the H_{011} mode with an inhomogeneous dielectric of this type.

(7.3.2) **Resonator-Action Theorem.**

The analysis is based upon the resonator-action theorem of Maclean.[30] The theorem, as stated by Maclean, is as follows: "In a loss-less electromagnetic resonator, the action of each mode, i.e. the product of the total energy and period, is invariant against an adiabatic deformation." For all practical purposes an adiabatic deformation is one in which the amount of deformation of the cavity per period of oscillation is small. From this theorem it follows that, if the electromagnetic energy of a resonator is increased from W_0 to $W_0 + \delta W$, the resonant frequency increases from f_0 to $f_0 + \delta f$, where

$$\frac{\delta W}{W_0} = \frac{\delta f}{f_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

(7.3.3) **Approximate Condition for Resonance.**

An approximate condition for resonance based upon the resonator action can be derived in the following manner:

Consider the motion of an electron distant z inside the cavity and anywhere within the annulus between r and $r + \delta r$, as shown in Fig. 14. The electron oscillates under the action of

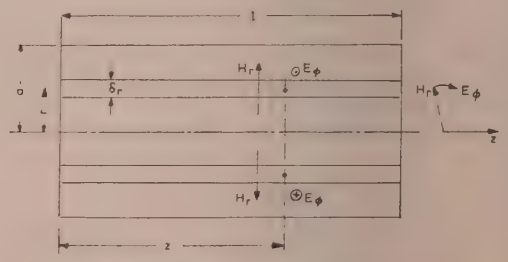


Fig. 14.—The field acting upon an electron.

the electric field, and this velocity is at right angles to the magnetic field $H_r(r)$. A longitudinal force is therefore exerted on the electron, and work has to be done on the electron to bring it up to the point z.

The total work done in filling up the region of the discharge with electrons is given by

$$W = \frac{1}{8}\frac{e^2}{\omega^2 m}\int_0^a N(r)E_\phi^2(r)2\pi r dr \quad . \quad . \quad . \quad (33)$$

where $N(r)$ is the electron density at radius r.

The total energy originally stored in the cavity is given by

$$W_0 = \int_{vol.\ of\ cavity} \left[\frac{1}{2}\epsilon_0\left(\frac{\hat{E}}{\sqrt{2}}\right)^2 + \frac{1}{2}\mu\left(\frac{\hat{H}}{\sqrt{2}}\right)^2\right] dv \quad . \quad . \quad (34)$$

and eqn. (32) becomes

$$\frac{W}{W_0} = \frac{\delta f}{f_0} = \left(\frac{k_3}{k_1}\right)^2 \frac{\int_0^a \Delta\epsilon_1'(r) r J_1^2(k_1 r) dr}{R^2 J_0^2(k_3 R)} \quad . \quad . \quad (35)$$

where $$\Delta\epsilon_1'(r) = \frac{e^2 N(r)}{\omega^2 m \epsilon_0}$$

It is convenient, and in this case it does not introduce much error, to write $J_1(k_1 r) = k_1 r/2$, and eqn. (35) leads to the simplified equation

$$\frac{\delta f}{f_0} = \frac{\int_0^a \frac{1}{4}\Delta\epsilon_1'(r) r^3 dr}{\left[\frac{1}{k_3} R J_0(k_3 R)\right]^2} \quad . \quad . \quad . \quad . \quad . \quad (36)$$

This equation does not define uniquely the function $\Delta\epsilon_1'(r)$ in terms of the measured frequency shift δf, but it does enable the values of δf for a given average electron density to be compared for different distributions of the electron density. For a uniform discharge, eqn. (36) becomes

$$\delta f = 0{\cdot}0069 f_0 \Delta\epsilon_1'$$

The curves of $\Delta\epsilon_1'$ versus δf deduced from this equation agree very satisfactorily with those deduced from eqn. (29), i.e. from the accurate solution of the two-media case, within the range $1{\cdot}00 \geqslant \epsilon_1' > 0{\cdot}9$.

Close agreement can be expected only for small values of $\Delta\epsilon_1'$, because in the approximate analysis the electric and magnetic fields were assumed to be unchanged by the presence of the electrons within the discharge.

Eqn. (36) shows that the change in the resonant frequency depends both upon the total number of electrons and their distribution within the discharge. To compare the effect of different distributions it is convenient to base the calculations upon an assumed average electron density, N_a, say.

The radial distribution of electron density deduced by Killian[19] from probe measurements follows quite closely the curve of the form

$$\hat{N}\left[1 - \frac{7}{8}\left(\frac{r}{a}\right)^2\right]$$

Table 4 shows the relation between $\Delta\epsilon_1'$ and δf for this and the three other electron-density distributions shown in Fig. 15.

When the electron density is calculated from the change in the relative permittivity of the discharge, the result obviously depends upon which distribution had been assumed in the theory relating change in resonant frequency to permittivity. For small changes of permittivity it is possible to deduce the value of $\Delta\epsilon_1'$ from the theory assuming uniform distribution and then multiply this quantity by the factor in Table 4. If the electron-density distribution is approximately parabolic but the change in the relative permittivity is calculated assuming it to be uniform, the calculated value of the average electron density is approximately two-thirds of the true value.

(7.3.4) Relation between Q-Factor and the Electron Collisional Frequency.

It has already been pointed out that the Q-factor of the cavity containing a discharge is a function of the collisional frequency of the electrons. An approximate analysis is given in this Section which leads to a simple relation between these quantities.

Assuming that the electric field within the cavity is unaltered by the discharge, the change of energy stored in the region occupied by the discharge is

$$\Delta W = \int_{vol.\ of\ discharge} \tfrac{1}{2}\epsilon_0(\epsilon_1' - 1)E^2 dv$$

By a well-known theorem the change in energy due to the change in electric field strength is equal to $-2\Delta W$, and so if the change in electric field strength, E, is taken into account the expression for ΔW becomes

$$\Delta W = \int_{vol.\ of\ discharge} \tfrac{1}{2}\epsilon_0 E^2(1 - \epsilon_1') dv \quad . \quad . \quad . \quad . \quad (37)$$

The power dissipated in the discharge is given by

$$P = \int_{vol.\ of\ discharge} \sigma E^2 dv$$

Table 4

	(a) Uniform	(b) Parabolic	(c) Killian type	(d) Quasi-parabolic
Distribution	$N(r) = \hat{N}$	$N(r) = \hat{N}\left[1 - \left(\frac{r}{a}\right)^2\right]$	$N(r) = \hat{N}\left[1 - \frac{7}{8}\left(\frac{r}{a}\right)^2\right]$	$0 < r < \frac{a}{2}$; $N(r) = \hat{N}$ $\frac{a}{2} < r < a$; $N(r) = \hat{N}\left(1 - \frac{r}{a}\right)$
$\hat{N}$ in terms of N_a	1	2	$\frac{16}{9}$	$\frac{12}{7}$
δf in terms of $\Delta\epsilon_1'$	$0{\cdot}0069 f_0 \Delta\epsilon_1'$	$0{\cdot}0046 f_0 \Delta\epsilon_1'$	$0{\cdot}0052 f_0 \Delta\epsilon_1'$	$0{\cdot}0046 j \Delta\epsilon_1'$
$\Delta\epsilon_1'$ in terms of δf	$145\frac{\delta f}{f_0}$	$216\frac{\delta f}{f_0}$	$194\frac{\delta f}{f_0}$	$216\frac{\delta f}{f_0}$
Distribution factor ..	$\frac{145}{145} = 1$	$\frac{216}{145} = 1{\cdot}49$	$\frac{194}{145} = 1{\cdot}34$	$\frac{216}{145} = 1{\cdot}49$

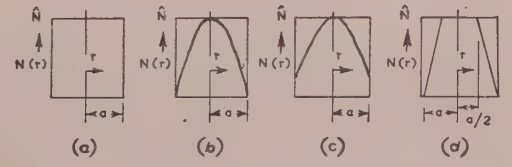


Fig. 15.—Electron-density distribution.

(a) Uniform.
(b) Parabolic.
(c) Killian type.
(d) Quasi-parabolic.

Eqns. (32) and (37) lead to the relation

$$W_0 = \frac{f}{\delta f}\int_{vol.\ of\ discharge} \tfrac{1}{2}\epsilon_0 E^2(1 - \epsilon_1')dv$$

The Q-factor of the cavity is defined by

$$Q = \frac{\omega \times \text{Energy stored}}{\text{Power loss}}$$

It is well known[31] that if the Q-factor of a cavity containing a lossy dielectric is Q_L (say) then

$$\frac{1}{Q_L} = \frac{1}{Q_0} + \frac{1}{Q_d}$$

where Q_0 is the Q-factor of the cavity with no discharge and

Q_d is the Q-factor of the lossy dielectric

i.e.
$$Q_d = \frac{\omega W_0}{P}$$

$$= \omega \frac{f}{\delta f}\frac{\int \tfrac{1}{2}\epsilon_0 E^2(1 - \epsilon_1')dv}{\int \sigma E^2 dv} \quad . \quad . \quad . \quad (38)$$

The integrations are carried out throughout the volume of the discharge. By comparing the expressions for σ and ϵ_1' [eqns. (3) and (4)], it is seen that eqn. (38) leads to

$$Q_d = \frac{f}{\delta f}\frac{\omega}{2\nu}$$

i.e.
$$\frac{1}{Q_L} - \frac{1}{Q_0} = \frac{f}{\delta f}\frac{\omega}{2\nu}$$

and therefore
$$\nu = \frac{\omega}{2}\frac{f}{\delta f}\frac{1}{Q_0}\left(\frac{Q_0}{Q_L} - 1\right) \quad . \quad . \quad . \quad (39)$$

This expression enables the value of ν to be calculated from the experimental results.

Eqn. (39) can be rearranged to give

$$Q_L = \frac{Q_0}{1 + Q_0\frac{2\nu}{\omega}\frac{\delta f}{f}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

This enables the Q-factor to be calculated directly from the measured change in the resonant frequency for an assumed value of ν.

621.313.044.6 : 621.3.017.31

Monograph No. 123
Mar. 1955

A NOTE ON THE SURFACE LOSS IN A LAMINATED POLE-FACE

By Professor G. W. CARTER, M.A., Member.

(*The paper was first received* 30*th October*, 1954, *and in revised form* 10*th January*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in March*, 1955.)

SUMMARY

Linear electromagnetic theory has been used to obtain a formula for the surface eddy-current loss in a laminated pole-face due to the flux ripples induced by the opposing slotted armature, without any restrictive assumption being made as to the thickness of the laminations. Earlier formulae have been examined in the light of this more general theory, and a formula derived by the author's father, the late Dr. F. W. Carter, has been found to be a good approximation over a range of conditions frequently encountered in practice.

LIST OF SYMBOLS

The M.K.S. system of units has been used.

$\boldsymbol{B}$ = Magnetic flux density, webers/m^2.
B_x, B_y, B_z = Components of $\boldsymbol{B}$.
B_1 = Peak value of the sinusoidal ripple in the normal flux density at the pole-face, webers/m^2.
d = Characteristic depth of penetration of flux and current in a given conducting material, m.
$\boldsymbol{E}$ = Electric force, volts/m.
f_r = Frequency of flux ripple, c/s.
h = Thickness of lamination, m.
h_c = Value of h around which F. W. Carter's formula is valid.
$\boldsymbol{H}$ = Magnetizing force, AT/m.
$\boldsymbol{J}$ = Current density, amp/m^2.
J_x, J_y, J_z = Components of $\boldsymbol{J}$.
$\hat{J}_{xn}$, $\hat{J}_{yn}$ = Maximum values of the nth space-components of J_x, J_y.
n = Order of harmonic in a Fourier expansion in the co-ordinate x.
P_e = Eddy-current loss per unit surface area, watts/m^2.
q = Wavelength of inducing flux ripple (slot pitch on opposing member), m.
v = Peripheral velocity of rotor, m/sec.
γ_n = Complex number describing attenuation and phase shift of B and J with increasing depth below surface, m^{-1}.
α_n, β_n = Real and imaginary parts of γ_n.
μ = Absolute permeability, henrys/m.
μ_0 = Absolute permeability of free space, $4\pi \times 10^{-7}$ henry/m.
ρ = Resistivity, ohm-m.

(1) INTRODUCTION

Formulae for the surface eddy-current loss in a pole-face, consequent upon the passage of a sinusoidal travelling wave of flux density, were derived by F. W. Carter.[1] Two cases were considered, namely the solid pole-shoe and the pole-shoe divided into very thin laminations; for the latter type the loss per unit area was found to be proportional to the square of the lamination thickness. More recently Gibbs,[2] evidently believing that the pole-shoe laminations used in practice are too thick for Carter's laminated-shoe formula to be applicable, derived a new formula in which the lamination thickness appears to the first power, and so gave a rational explanation of experimental results obtained by Spooner and Kinnard.[3] Walker[4] modified Gibbs's formula to make it suitable for surface-loss calculations in induction motors.

Applied to a particular design of small induction motor, Gibbs's and Walker's formulae were found to give losses which exceeded the value obtained from Carter's laminated-shoe formula by ratios of 30 or 40 to 1. It therefore seemed desirable to calculate the surface eddy-current loss by classical theory without making any assumption of thin laminations, in the hope of throwing light upon the proper range of validity for each formula.

(2) CALCULATION OF SURFACE EDDY-CURRENT LOSS IN A LAMINATED POLE-FACE

The pole-face will be considered as a plane surface. The origin of Cartesian co-ordinates is taken in this surface at the centre of a lamination, with the x-axis parallel to the slots in the opposing member, the y-axis parallel to the direction of motion of the opposing member, and the z-axis inward to the pole. The material of the pole-face is considered to have constant absolute permeability μ and resistivity ρ. The opposing member, moving across the pole-face with relative velocity v, sets up to the surface of the pole a perpendicular flux-density given by

$$B_{z0} = B_1 \cos\frac{2\pi}{q}(y - vt) \quad . \quad . \quad . \quad . \quad (1)$$

a gliding flux-ripple, which is superimposed upon a uniform flux that is without significance in the calculation of eddy currents.

The electric force $\boldsymbol{E}$, magnetic field vectors $\boldsymbol{B}$, $\boldsymbol{H}$, and induced current density $\boldsymbol{J}$, set up within the pole, are subject to the equations

$$\text{curl } \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$\text{curl } \boldsymbol{H} = \boldsymbol{J} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$\boldsymbol{E} = \rho \boldsymbol{J} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

$$\boldsymbol{B} = \mu \boldsymbol{H} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

from which may be deduced that each component of any of the vectors satisfies an equation of the form

$$\nabla^2 J - \frac{\mu}{\rho}\frac{\partial J}{\partial t} = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The applied surface flux-density, eqn. (1), is the real part of $B_1 \varepsilon^{2\pi j(y-vt)/q}$. The planes $x = \pm h/2$ form the surfaces of a single lamination, over the outer edge of which the flux-density equation, (1), may be expressed as a Fourier series:

$$B_{z0} = \frac{4}{\pi}B_1\left(\cos\frac{\pi x}{h} - \frac{1}{3}\cos\frac{3\pi x}{h} + \frac{1}{5}\cos\frac{5\pi x}{h} - \ldots\right)\varepsilon^{2\pi j(y-vt)/q}$$

$$= \frac{4}{\pi}B_1\varepsilon^{2\pi j(y-vt)/q}\sum_{n\,\text{odd}}\frac{(-1)^{\frac{n-1}{2}}}{n}\cos\frac{n\pi x}{h} \quad . \quad (7)$$

Correspondence on Monographs is invited for consideration with a view to publication.

Professor Carter is Professor of Electrical Engineering, Leeds University.

The following set of components of $\boldsymbol{B}$ and $\boldsymbol{J}$ will be found to satisfy eqn. (6), to give no outflow of current from the surfaces of a lamination, and to be related among themselves in accordance with eqns. (2)–(5); furthermore, when z is equated to zero in B_z, the expression reduces to eqn. (7):

$$B_x = \frac{4}{\pi^2} B_1 h \varepsilon^{2\pi j(y-vt)/q} \sum_{n\,\text{odd}} \frac{(-1)^{\frac{n-1}{2}} \gamma_n \varepsilon^{-\gamma_n z}}{n^2 + (2h/q)^2} \sin\frac{n\pi x}{h} \quad . \quad (8)$$

$$B_y = -\frac{8j}{\pi^2} \frac{B_1 h^2}{q} \varepsilon^{2\pi j(y-vt)/q} \sum_{n\,\text{odd}} \frac{(-1)^{\frac{n-1}{2}} \gamma_n \varepsilon^{-\gamma_n z}}{n[n^2 + (2h/q)^2]} \cos\frac{n\pi x}{h} \quad (9)$$

$$B_z = \frac{4}{\pi} B_1 \varepsilon^{2\pi j(y-vt)/q} \sum_{n\,\text{odd}} \frac{(-1)^{\frac{n-1}{2}} \varepsilon^{-\gamma_n z}}{n} \cos\frac{n\pi x}{h} \quad . \quad . \quad (10)$$

$$J_x = -\frac{16}{\pi} \frac{B_1 v h^2}{\rho q^2} \varepsilon^{2\pi j(y-vt)/q} \sum_{n\,\text{odd}} \frac{(-1)^{\frac{n-1}{2}} \varepsilon^{-\gamma_n z}}{n[n^2 + (2h/q)^2]} \cos\frac{n\pi x}{h} \quad (11)$$

$$J_y = \frac{8j}{\pi} \frac{B_1 v h}{\rho q} \varepsilon^{2\pi j(y-vt)/q} \sum_{n\,\text{odd}} \frac{(-1)^{\frac{n-1}{2}} \varepsilon^{-\gamma_n z}}{n^2 + (2h/q)^2} \sin\frac{n\pi x}{h} \quad . \quad (12)$$

$$J_z = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

In these equations,

$$\gamma_n^2 = \frac{n^2\pi^2}{h^2} + \frac{4\pi^2}{q^2} - j\frac{2\pi\mu v}{\rho q} \quad . \quad . \quad . \quad (14)$$

In terms of real quantities, eqns. (11) and (12) are equivalent to

$$\left.\begin{aligned} J_x &= \sum_{n\,\text{odd}} \hat{J}_{xn} \cos\left[\frac{2\pi}{q}(y - vt) - \beta_n z\right] \\ J_y &= \sum_{n\,\text{odd}} \hat{J}_{yn} \sin\left[\frac{2\pi}{q}(y - vt) - \beta_n z\right] \end{aligned}\right\} \quad . \quad . \quad (15)$$

where $$\hat{J}_{xn} = -\frac{16}{\pi} \frac{B_1 v h^2}{\rho q^2} \frac{(-1)^{\frac{n-1}{2}} \varepsilon^{-\alpha_n z}}{n[n^2 + (2h/q)^2]} \cos\frac{n\pi x}{h} \quad . \quad (16)$$

$$\hat{J}_{yn} = -\frac{8}{\pi} \frac{B_1 v h}{\rho q} \frac{(-1)^{\frac{n-1}{2}} \varepsilon^{-\alpha_n z}}{n^2 + (2h/q)^2} \sin\frac{n\pi x}{h} \quad . \quad . \quad (17)$$

and $$\gamma_n = \alpha_n + j\beta_n \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

The mean rate of loss per unit area of pole-face is then given by

$$P_e = \frac{\rho}{2h} \sum_{n\,\text{odd}} \int_{-\frac{h}{2}}^{\frac{h}{2}} \int_0^\infty (\hat{J}_{xn}^2 + \hat{J}_{yn}^2) dx dz \quad . \quad . \quad (19)$$

which reduces to

$$P_e = \frac{8}{\pi^2} \frac{B_1^2 f_r^2 h^2}{\rho} \sum_{n\,\text{odd}} \frac{1}{\alpha_n n^2 [n^2 + (2h/q)^2]} \quad . \quad . \quad (20)$$

where $f_r = v/q$, the frequency of the inducing flux-ripple. This is the general loss formula, applicable to laminations of all thicknesses.

(4) DISCUSSION OF THE SURFACE-LOSS FORMULA

The quantity α_n in eqn. (20) is given by eqn. (14) and (18), which may be written

$$\alpha_n + j\beta_n = \left[\left(\frac{n^2\pi^2}{h^2} + \frac{4\pi^2}{q^2}\right) - \frac{j}{d^2}\right]^{1/2} \quad . \quad . \quad (21)$$

where $$d = \left(\frac{\rho}{2\pi f_r \mu}\right)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

d is the characteristic "depth of penetration"[5] of flux into a solid block of this material at the frequency f_r.* Even at the supply frequency, the depth of penetration is likely to be smaller than the slot pitch q; at the considerably higher ripple frequency we shall expect to find that $(2\pi/q)^2$ is much less than $1/d^2$ in most practical cases. The middle term on the right-hand side of eqn. (21) is therefore not very important; but the first term can exceed the third if the laminations are thin enough.

For very thick laminations, the first term becomes negligible; neglecting the second also, we obtain

$$\alpha_n = -\beta_n = \frac{1}{d\sqrt{2}} \text{ (approximately)} \quad . \quad . \quad (23)$$

The loss per unit area is then given by

$$P_e = \frac{2\sqrt{2}}{\pi^2} \frac{B_1^2 f_r^2 q^2 d}{\rho} \sum_{n\,\text{odd}} \frac{1}{n^2}$$

$$= \frac{1}{4\pi^{1/2}} \frac{B_1^2 f_r^{3/2} q^2}{\mu^{1/2}\rho^{1/2}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

since $\Sigma 1/n^2$ (odd terms) $= \pi^2/8$. Eqn. (24) is essentially the same as F. W. Carter's formula for the unlaminated pole [eqn. (20) of his paper].

For very thin laminations the first term on the right-hand side of eqn. (21) predominates even when $n = 1$. We now have

$$\alpha_n = \frac{n\pi}{h}, \quad \beta_n = 0 \text{ (approximately)} \quad . \quad . \quad (25)$$

so that $$P_e = \frac{8}{\pi^3} \frac{B_1^2 f_r^2 h^3}{\rho} \sum_{n\,\text{odd}} \frac{1}{n^5}$$

$$= 0{\cdot}2592 \frac{B_1^2 f_r^2 h^3}{\rho} \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

This bears no resemblance to F. W. Carter's laminated-pole formula [eqn. (23) of his paper]; in our notation, this would be

$$P_e = \frac{\pi^{3/2}}{12} \frac{B_1^2 f_r^{3/2} h^2}{\mu^{1/2}\rho^{1/2}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

To discern the significance of this expression we must observe that Dr. Carter makes two assumptions: that the x-component of current density is negligible compared with the y-component, and that the rate at which the induced currents decline with increasing z—a decline characterized by our coefficients α_n—is given by an equation [his eqn. (22)] which is essentially similar to our eqn. (23). It follows that F. W. Carter's laminated-pole formula is applicable, not to the thinnest laminations, but to laminations which are thick enough to ensure that the final term in eqn. (21) shall predominate, yet thin enough for the transverse component of current density to be regarded as negligible by comparison with the longitudinal component. Will these conditions be simultaneously satisfied in practice? It is probable that they will, if the ratio of q (the wavelength of the inducing flux-wave) to d (the depth of penetration) is sufficiently great. Since the validity of the formula depends on the simultaneous evanescence of $(2h/q)^2$ and $(\pi d/h)^2$, it is likely to be found to be most accurate when the lamination thickness is near the value

$$h_c = \left(\frac{\pi}{2} dq\right)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad (28)$$

Eqn. (27) may be derived from eqn. (20) by substituting for α_n the value given in eqn. (23) and treating $(2h/q)^2$ as negligible.

* Some writers use the same term for a depth $\sqrt{2}$ times as great as this.

Both of F. W. Carter's formulae are therefore shown to be special cases of the present more general theory.

(5) NUMERICAL DISCUSSION

The infinite series appearing in the general loss formula, (20), is a function of three variables, h, q and d, and therefore does not lend itself easily to graphical representation. A clearer view is obtained by discussing a particular case. The various formulae will be applied to the calculation of the stator surface loss in a particular type of small induction motor, and to investigating the effect of varying the lamination thickness while keeping all the other parameters constant.

The motor was a 4-pole 50c/s machine having 43 rotor slots; at synchronous speed the frequency of the flux oscillation at the stator surface, f_r, was therefore 1 075c/s. The rotor slot pitch, q, was 0·874cm. The stator iron had a resistivity, ρ, of 35×10^{-8}ohm-m; for the relative permeability, μ/μ_0, three values were assumed, namely 1 000, 2 000 and 4 000.

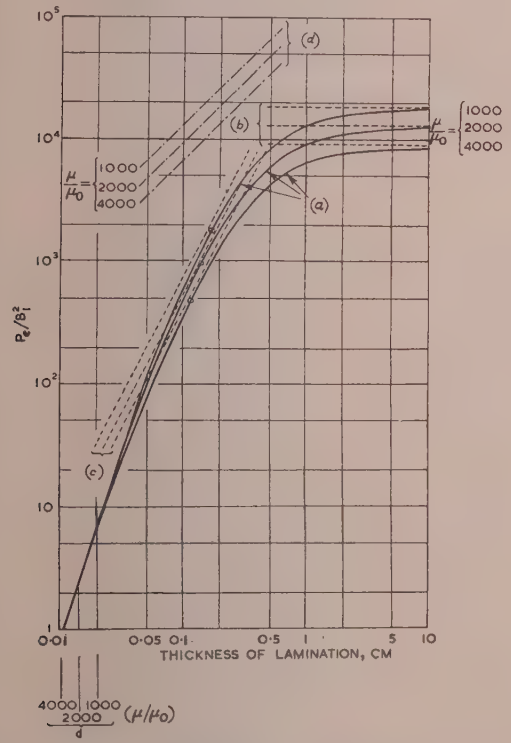


Fig. 1.—Application of eqn. (20) to the experimental machine.

(*a*) Author's formula.
(*b*) F. W. Carter's solid-pole formula.
(*c*) F. W. Carter's laminated-pole formula.
(*d*) Gibbs's formula.

The ringed points indicate centres of validity zones for F. W. Carter's formula [eqn. (28)].
The wavelength of the inducing flux-ripple was 0·874 cm.

From these data, the curves in Fig. 1 have been drawn, representing eqn. (20) in its application to this machine. For the thinnest laminations the surface loss is proportional to the cube of the lamination thickness, h, and is independent of the permeability; this is because the depth of penetration of the flux is here controlled by the lamination thickness rather than by the properties of the iron. At somewhat greater lamination thicknesses a zone is found in which F. W. Carter's laminated-pole formula (making the loss proportional to h^2) is quite a good approximation. For very thick laminations the loss tends towards the value given by F. W. Carter's solid-pole formula.

Investigating the surface loss in laminated pole-shoes, Gibbs* concluded that the thick laminations there employed would fall into a region intermediate between the zones of application of Dr. Carter's two formulae. For this region he devised a formula making the loss proportional to h; in the units and notation of the present paper, this formula is

$$P_e = (4{\cdot}03 \times 10^4) B_1^2 f_r^{3/2} q h \quad . \quad . \quad . \quad . \quad (29)$$

This is based, however, on a particular value of μ, namely that which makes $(\rho\mu/\mu_0)^{1/2}$ equal to 0·02. In general, Gibbs's formula would be

$$P_e = 0{\cdot}904 \frac{B_1^2 f_r^{3/2} q h}{\mu^{1/2} \rho^{1/2}} \quad . \quad . \quad . \quad . \quad . \quad (30)$$

Walker,† basing his opinion partly on experimental evidence, advocates the use of this formula in a modified form for induction motors, in spite of the much thinner laminations which would be used. Walker's formula, however, is not quite so easy to interpret as Gibbs's, so the latter has been used for Fig. 1, allowance being made for various values of permeability by using eqn. (30).

It will be seen that, for the conditions obtaining in this machine, the Gibbs formula is not a good approximation. Even for that part of the curve where the loss varies approximately as the first power of the lamination thickness, Gibbs's formula gives a loss which is too high by a factor of eight. For laminations of the thickness which is customary in induction motors—in the motor referred to, the thickness was 0·066cm—the discrepancy is greater still; on the other hand, Dr. Carter's laminated-pole formula gives a value quite near to that given by the author's formula, eqn. (20).

Also indicated in Fig. 1 are the characteristic depths of penetration, d, for each value of permeability. From these, by the use of eqn. (28), we may deduce the approximate centres ($h = h_c$) of the ranges of validity of F. W. Carter's laminated-pole formula. These are indicated by ringed points, and will be seen to be very near to the points of closest approach between Dr. Carter's curves and the author's. Eqn. (28) may therefore be used as a rapid means of determining whether Dr. Carter's formula is likely to be valid. For example, if we consider the alternator described in Appendix 15.4 of Gibbs's paper, for which $q = 2{\cdot}8$cm, $f_r = 1\,125$c/s, and suppose it to have laminated poles of mild steel for which $\rho = 20 \times 10^{-8}$ohm-m, $\mu/\mu_0 = 2\,000$, we shall find that $h_c = 0{\cdot}22$cm. This lies right in the middle of the range of thicknesses commonly used in practice.

(6) CONCLUSION AND ACKNOWLEDGMENTS

It must not be forgotten that linear electromagnetic theory can only give approximate results in a problem of this kind. Experimental results which conflict with the formulae here given, such as Spooner and Kinnard's demonstration‡ of a linear variation of loss with lamination thickness down to a thickness of 1mm, may well be associated with a partial failure of the linear theory, e.g. with a variation of effective permeability with varying lamination thickness. The present paper has merely compared several formulae which are all based on linear electromagnetic theory, and has shown that some are more securely based than others.

The author wishes to record his thanks to his colleague, Mr. J. Hanania, for preparing the diagram; also to Dr. W. J. Gibbs, for data supplied by correspondence.

* *Op. cit.*, p. 7. † *Op. cit.*, p. 603. ‡ *Op. cit.*, Fig. 10.

(7) REFERENCES

(1) CARTER, F. W.: "Pole-Face Losses," *Journal I.E.E.*, 1916, **54**, p. 168.
(2) GIBBS, W. J.: "Tooth-Ripple Losses in Unwound Pole-Shoes," *ibid.*, 1947, **94**, Part II, p. 2.
(3) SPOONER, T., and KINNARD, I. F.: "Surface Iron Losses with Reference to Laminated Materials," *Transactions of the American I.E.E.*, 1924, **43**, p. 262.
(4) WALKER, J. H.: "A Theory of Induction-Motor Surface Losses," *Journal I.E.E.*, 1948, **95**, Part II, p. 597.
(5) ABRAHAM, M., and BECKER, R.: "Classical Electricity and Magnetism" (Blackie, London, 1932), p. 199.

621.3.015.54:621.3.015.532

Monograph No. 124 M
Apr. 1955

THE INITIATION MECHANISM OF LONG SPARKS IN POINT-PLANE GAPS

By R. F. SAXE, Ph.D., and Prof. J. M. MEEK, D.Eng., Member.

(*The paper was first received* 4*th June*, 1954, *in revised form* 18*th October*, 1954, *and in final form* 27*th January*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in April*, 1955.)

SUMMARY

Investigations have been made of the optical and electrical characteristics of the corona discharges and leader strokes preceding the impulse breakdown of positive-point/negative-plane gaps in air at atmospheric pressure. The results show that initially a corona discharge forms in the highly stressed region round the pointed electrode; this corona gives rise to a short-duration (~0·4 microsec) pulse of light emission in the gap, accompanied by a similar pulse of current in the circuit. The size and shape of the light-emitting volume in this corona discharge have been studied. The next stage in the growth of the spark is the development of the leader stroke from the corona. The leader stroke grows across the gap in a predictable manner, and the current flowing in the circuit during the growth depends on the velocity of the leader stroke. The size and configuration of the light-emitting volume of the leader stroke are similar to those of the corona. The characteristics of the corona and leader stroke in air at reduced pressures and in oxygen, nitrogen and hydrogen are also briefly described.

(1) INTRODUCTION

A number of investigations have been made in various laboratories to observe the mechanism of growth of spark discharges in long gaps. This introduction includes only a brief summary of the main features of the results of this work, with emphasis given to those which bear most closely on the present investigation. For a fuller account of the earlier work, up to 1952, the reader is referred to the recent book by Meek and Craggs.[1]

The mechanism of growth of lightning discharges has been investigated extensively by Schonland and his colleagues,[2–5] who have used mechanical-scanning camera techniques to photograph the visual growth. Their results show that each lightning discharge is preceded by a streamer type of process, known as a leader stroke, which grows rapidly from cloud to ground. When the leader stroke reaches ground the main stroke begins to travel up the ionized path which has been established by the leader stroke. The high currents associated with lightning occur during the period of the main stroke.

Allibone and Meek[6] have shown that the mechanism of growth of sparks across long gaps subjected to impulse voltages in the laboratory is essentially similar to that of lightning. Photographs, taken with a mechanical-scanning camera, show that for a gap between a positive point and a negative plane the discharge begins at the point with the formation of streamers. If the voltage applied to the gap is sufficiently high, the whole streamer process, or leader stroke, advances from the point to the plane. The leader stroke is recorded on the photographs as an intense filamentary channel accompanied by a less bright, more diffuse, "shower" of discharge. When the leader stroke has crossed the gap the main stroke occurs and follows in detail the tortuous path that has been traced out by the leader stroke. With the occurrence of the main stroke the final spark channel is formed. Later mechanical-scanning camera studies by Komelkov[7–12] and by other workers[13–17] give results in substantial agreement with those of Allibone and Meek.

Other investigations of the growth of sparks in long gaps have been made by removing the impulse voltage applied to the gap before the leader stroke has formed completely across the gap. Photographs can then be obtained with a stationary camera of the various stages of growth of the leader stroke preceding the formation of the main channel. In the most recent studies of this type, by Norinder and his colleagues,[18–22] the visual growth of discharges has been recorded in great detail, with results that supplement the data given by the earlier mechanical-scanning camera photographs.

The object of the experiments described here has been to extend the results of the earlier investigations and, by the use of a new measuring technique, to record features of the discharges that are not readily detectable by other methods. For example, in the mechanical-scanning camera studies of sparks occurring in circuits of low series resistance, it is difficult to distinguish the leader stroke from the main stroke because of the halation from the very bright main stroke and also because of the small separation on the negative between the images of leader stroke and main stroke. Further, the spatial distribution of the discharge can also confuse the details of the time relationship and introduce inaccuracies even in cases when the scanning speed is high enough to separate clearly the leader stroke from the main stroke. In order to overcome these difficulties a photo-electric technique for recording the visual growth of leader strokes has been developed.[23,24] This technique has been used in the experiments described in the paper, with results which give a higher degree of temporal resolution than was obtainable in the earlier mechanical-scanning camera studies. The photo-electric records have also been supplemented by measurements of the current flowing during the growth of leader strokes—a factor on which few data have hitherto been obtained.

The technique achieves a greater temporal resolution than in earlier investigations and also has the merit of a very much greater sensitivity to light than is obtained by a mechanical-scanning camera.

(2) EXPERIMENTAL TECHNIQUES

(2.1) Measurement of Leader-Stroke Currents

The discharge gap used throughout the experiments to be described consists of a point anode and a plane cathode. The gap is connected through a series resistor to an impulse voltage generator, with an output capacitance of 0·025μF and a maximum output voltage of 1 000 kV. In the earlier experiments by Allibone and Meek the pointed electrode is at the high potential, and the measurement of the current flow during the initiation process of the spark is complicated by the capacitive currents which flow to the earthed plane electrode. These capacitive currents are almost entirely eliminated in the present experiments by applying the impulse voltage to the plane electrode and by measuring the current flow in the earthy pointed electrode.

The current-measuring circuit used is as follows. The pointed

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Saxe is in the Electrical Engineering Department, Queen Mary College, University of London, and Professor Meek is Professor of Electrical Engineering, University of Liverpool.

The paper is based partly on E.R.A. Reports Ref. L/T183 and L/T188.

electrode is connected to earth through a resistor R_1, the value of which is so chosen that the voltage developed across it by the current which flows during the development of the leader stroke is suitable for direct display on a cathode-ray tube (*c.* 100 volts). Some protection of the cathode-ray tube must be ensured, however, when the much heavier current in the main stroke flows through R_1 with the possible generation of a voltage higher than that which the Y-plates of the cathode-ray tube are designed to withstand. This protection is provided by shunting R_1 with a non-linear resistor (Metrosil) whose impedance is high compared with R_1 during the leader-stroke period but which falls to a value low compared with R_1 when the main-stroke current flows. Owing to the high dielectric constant of Metrosil, the shunt capacitance placed in parallel with R_1 is appreciable, and as a result the time-constant of the recording system is of the order of 10^{-8} sec.

A typical oscillogram showing the current flowing in the pointed electrode as a function of time during the development of the leader stroke is shown in Fig. 1. The average values of

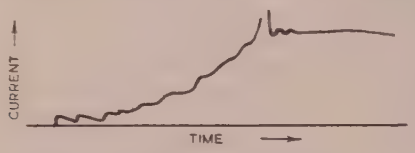


Fig. 1.—Current flow due to leader stroke.

not fewer than ten of these current/time curves were used in order to determine the curves given in Figs. 7 and 8. The current/time curves varied both in total duration and also in the current at any given fraction of the total duration due to random variations in the individual leader strokes. In order to obtain a mean current/time curve for any given set of curves, the individual time-scales were expanded or contracted to fit the mean duration of the set, and then the currents at any given time were averaged.

Owing to the considerable statistical fluctuations encountered in the phenomena investigated, no attempt was made to obtain great accuracy in the measurements. In most cases the accuracy obtained is probably about 5%.

(2.2) Measurement of the Rate of Growth of Leader Strokes

The rate of growth of the leader stroke has been studied by measurement of the times at which it reaches certain points in its travel. From these observations distance/time curves may be drawn whose slope at any point gives the velocity of the leader stroke.

The time at which the leader stroke reaches any given point in its travel is determined in the following manner. A lens forms an image of the spark-gap on an opaque screen in which there is a narrow slit perpendicular to the direction of movement of the leader stroke. When the luminous part of the leader stroke reaches a point such that its image is focused on the slit, light from it passes through the slit and falls upon the photomultiplier. The resultant electrical pulse emitted by the photomultiplier is amplified and displayed on a cathode-ray oscillograph. When the leader stroke has completed its movement across the gap, the main stroke occurs, and as this emits much more light than the leader stroke, it usually causes a break in the oscillograph trace, followed by a flat top due to saturation of the amplifier. A typical record is shown in Fig. 2.

As in the case of the current/time curves, allowance must be made for variations in the time-intervals between the application of the voltage and the occurrence of the main strokes in any

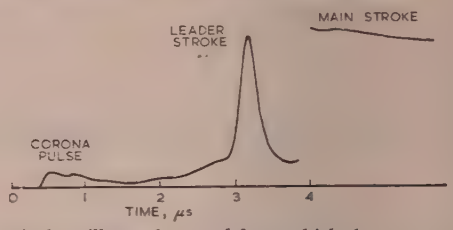


Fig. 2.—Typical oscillograph record from which the movement of the leader stroke is studied.

given set of records taken under the same conditions. This has been done by expansion or contraction of the individual time-scales to make them coincide with the average.

(3) EXPERIMENTAL RESULTS

(3.1) The Leader Stroke

Curves relating the distance that the brightly luminous part of the leader stroke has travelled with the time that has elapsed since the application of the voltage wave are shown in Fig. 3

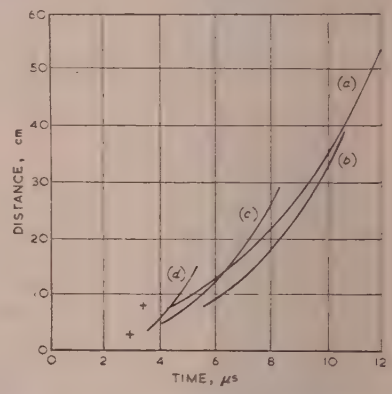


Fig. 3.—Effect of gap length on distance/time curves.

Series resistance, 23 kilohms. Gap size: (*a*) 55·4cm, (*b*) 40cm, (*c*) 30cm, (*d*) 16·3cm, + 8cm.

for gap lengths of 55·4, 40, 30, 16·3 and 8cm, with a series resistance of 23 000 ohms. The curves present the same general appearance; the leader stroke travels comparatively slowly at first and increases in velocity as it approaches the plane. Fig. 4

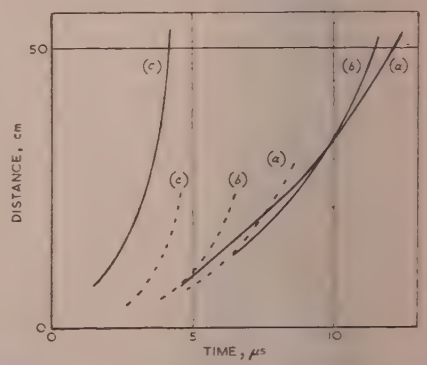


Fig. 4.—Effect of series resistance on distance/time curves for two values of gap length.

———— 55·4cm gap length.
– – – – 30cm gap length.

(*a*) 23-kilohm series resistance.
(*b*) 6·1-kilohm series resistance.
(*c*) 750-ohm series resistance.

shows the effect on the distance/time curve of a reduction in the series resistance; the distance at which each curve stops is the length of the gap for which the curve was obtained.

While the time taken for the leader stroke to propagate across the gap decreases with a reduction in either the gap length or the series resistance, the general aspect of the distance/time curve is relatively unaffected. This is made clear by an inspection of Figs. 5(*a*) and 5(*b*).

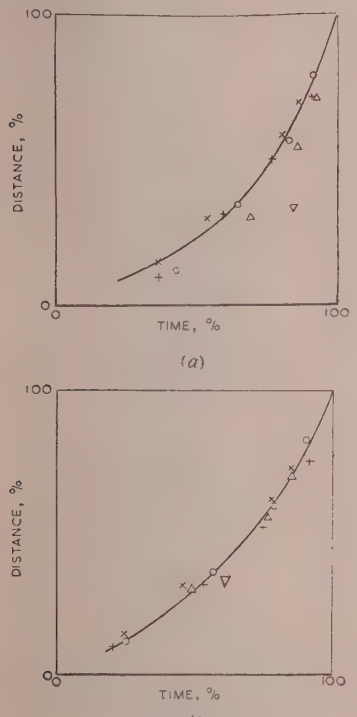


Fig. 5(a) and Fig. 5(b).—Percentage-distance/percentage-time curves for the movement of the leader stroke across the gap.

× 55·4cm gap.
\+ 40cm gap.
○ 30cm gap.
△ 16·3cm gap.
▽ 8cm gap.

In Fig. 5(*a*) curves are given to show the percentage of the gap length which the leader stroke has traversed at any given instant as a function of the time taken by the leader stroke to travel that distance, this time being expressed as a percentage of the time interval between the application of the voltage wave and the occurrence of the main stroke. The curves refer to the five gap lengths quoted above with a series resistance of 23 000 ohms. There is no appreciable difference between the curves for the three longer gaps, but there seems to be a tendency for the curves for the 16·3 cm and 8 cm gaps to lie to the right. This may be accounted for if it is assumed that the leader stroke does not begin to grow at the instant of application of the voltage wave, since at this instant, or immediately afterwards, the voltage is very low and the initiation mechanism may not have commenced. As the total time interval between the application of the voltage wave and the occurrence of the main stroke for the three longer gaps is large compared with that for the two smaller gaps, any correction due to the assumption above will leave the curves for the three longer gaps relatively unaffected but will cause appreciable changes in the curves for the two shorter gaps. Therefore, the curves for the longer gaps have been left uncorrected and those for the 16·3 cm and 8 cm gaps have been shifted to fit by the subtraction of the time interval from the total time of growth. It is found that the two figures for this time interval are 2·1 and 2·2 microsec respectively. All five curves, replotted on the assumption that the leader stroke commences to grow only after a delay of 2·1 microsec from the application of the voltage wave, are shown in Fig. 5(*b*). The agreement is within the limits of experimental error. The voltage wave for this particular experimental arrangement rises to its maximum value in approximately 2·1 microsec.

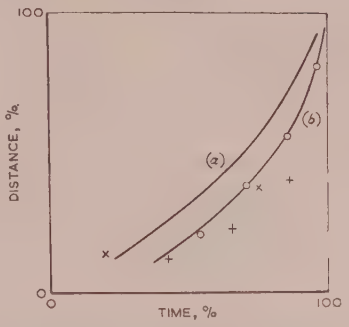


Fig. 6.—Effect of the series resistance on the percentage-distance/percentage-time curve.

(*a*) 23-kilohm series resistance
(*b*) 6·1-kilohm series resistance } 55·4cm gap
× 750-ohm series resistance
(*a*) 23-kilohm series resistance 30cm gap
\+ 6·1-kilohm series resistance
○ 750-ohm series resistance } 25cm gap

The curves in Fig. 6 show the effect of reduction of the series resistance on the percentage-distance/percentage-time curve. The curves are corrected for the effect of the time-lag between the instant of application of the impulse voltage and that of commencement of the leader stroke; the correction is assumed to be that which applies to the series resistance of 23 000 ohms, namely 1·3 times the value of CR where $C = 70\,\mu\mu$F (the approximate capacitance to earth of the high-voltage plane). The correction is negligible for the curves pertaining to a series resistance of 750 ohms. It will be seen from Fig. 6 that the effect of decreasing the series resistance is to decrease the slope of the curve during the initial period and to increase it subsequently.

Curves showing the variation of current in the pointed electrode with time for all five gap lengths measured above, and for a series resistance of 23 000 ohms, are shown in Fig. 7. By averaging several oscillograms to obtain each of the curves shown, a smooth curve is obtained, free from the small random irregularities which are present on the individual oscillograms. The rate of increase of current with time increases steadily until the main stroke occurs. The final current flowing just before the main stroke's occurrence increases with gap length, and is approximately 2·5 amp for the 8 cm gap and 15·5 amp for the 55·4 cm gap. The effect of reducing the series resistance to 6 100 ohms and 750 ohms is shown for a gap length of 55·4 cm in Fig. 8. The final current, measured in the pointed electrode just before the main stroke, increases considerably as the series resistance is lowered. From the current/time and distance/time curves it is possible to plot a curve connecting the charge that has flowed through the current shunt and the distance that the streamer has grown. Curves of this type for the five gap lengths measured above and for a series resistance of 23 000 ohms are shown in Fig. 9. These curves are approximately linear and have roughly the same slope, 0·88 microcoulomb/cm.

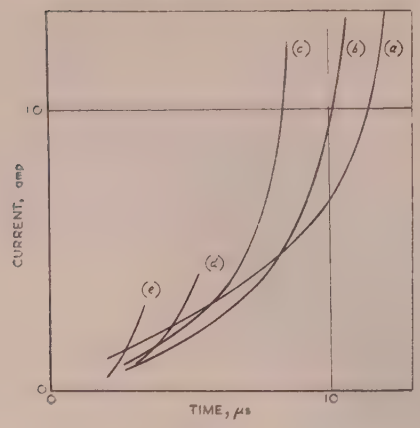


Fig. 7.—Effect of the gap length on the current in the leader stroke.

Series resistance, 23 kilohms.
(*a*) 55·4cm gap.
(*b*) 40cm gap.
(*c*) 30cm gap.
(*d*) 16·3cm gap.
(*e*) 8cm gap.

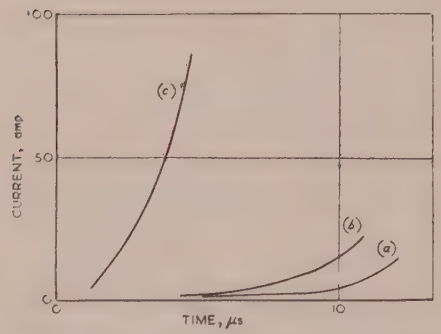


Fig. 8.—Effect of the series resistance on the current in the leader stroke.

(*a*) 23-kilohm series resistance.
(*b*) 6·1-kilohm series resistance.
(*c*) 750-kilohm series resistance.

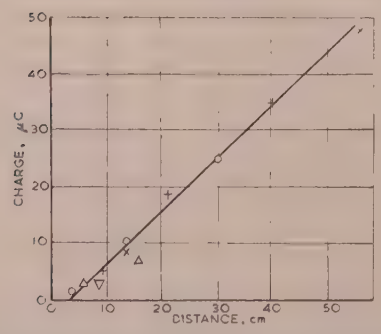


Fig. 9.—The charge flow caused by the movement of the leader stroke.

Series resistance, 23 kilohms.
× 55·4cm gap.
\+ 40cm gap.
○ 30cm gap
△ 16·3cm gap.
▽ 8cm gap.

The effect of applying a voltage greater than the minimum breakdown voltage (defined as the voltage at which breakdown occurred on approximately 90% of the applied impulses) has been investigated by reducing the gap length without altering the applied voltage. Curve (*a*), Fig. 10(*a*), shows the current/time

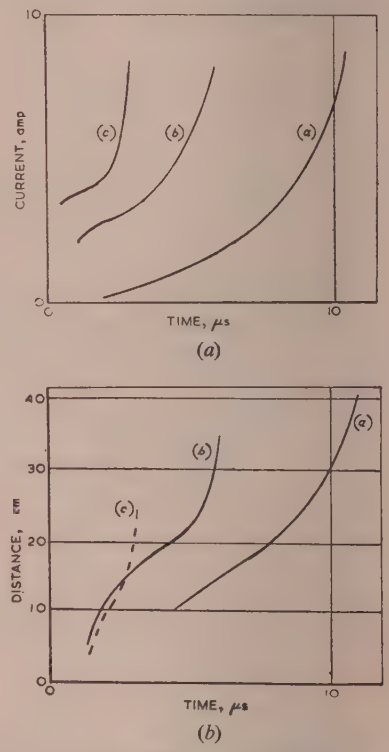


Fig. 10(a) and Fig. 10(b).—Effect on the current in, and the movement of, the leader stroke of reducing the gap length at fixed voltage.

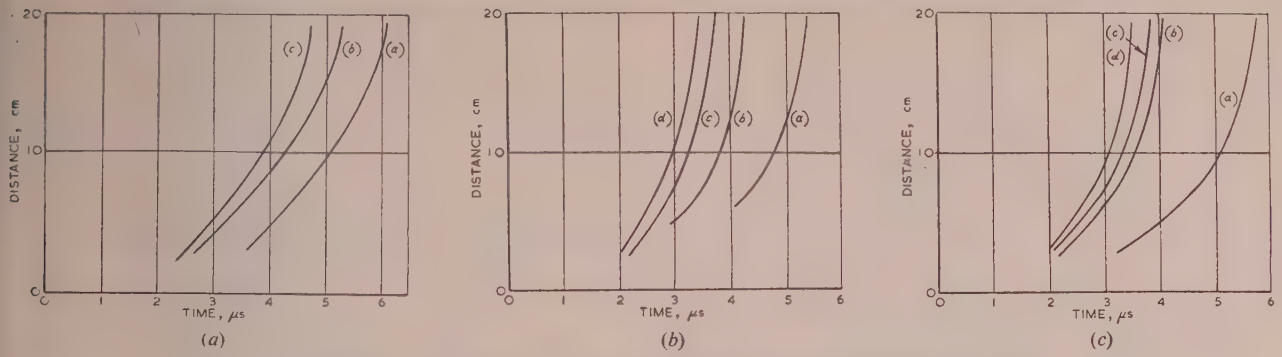
(*a*) 42·5cm gap.
(*b*) 35cm gap.
(*c*) 25cm gap.

curve for a gap length of 42·5 cm at its minimum breakdown voltage, while curves (*b*) and (*c*) show the effect of reducing this gap length to 35 cm and 25 cm respectively. The distance/time curves for the three cases are shown in Fig. 10(*b*).

The characteristics of the leader stroke as a function of the gas pressure have also been determined. The tests were conducted between a pointed electrode and a plane of diameter approximately 30 cm in a Perspex cylinder of diameter approximately 38 cm. For the gap lengths used, the spark took place always near to the centre of the plane. The distance/time curves and current/time curves for a 20 cm gap are shown in Figs. 11 and 12 for pressures of 760, 593, 435 and 277 mm Hg and with series resistances of 1 850, 3 850 and 23 000 ohms. The percentage-distance/percentage-time curves for these are shown in Fig. 13.

Curves relating the distance that the leader stroke has travelled across the gap with the charge that has flowed through the pointed electrode are shown in Fig. 14. These curves tend to be approximately linear and to have approximately the same slope, but to be displaced with respect to each other.

During experiments with the photomultiplier to obtain distance/time curves, an effect was observed which can be seen in Fig. 2. If the leader stroke consists of a point source, the pulse of light shown on the oscillogram of Fig. 2, and designated "leader stroke," would be of duration of the order of 10^{-7} sec.

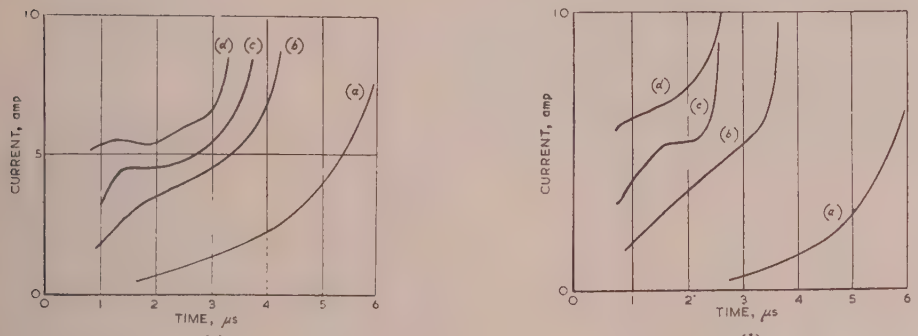


(a) (b) (c)

Fig. 11.—Effect of the reduction of gas pressure on the movement of the leader stroke.

(*a*) 23-kilohm series resistance.
(*b*) 3 850-ohm series resistance.
(*c*) 1 850-ohm series resistance.

Curve (*a*) 760mm Hg pressure.
Curve (*b*) 593mm Hg pressure.
Curve (*c*) 435mm Hg pressure.
Curve (*d*) 227mm Hg pressure.



(a) (b)

Fig. 12.—Effect of the reduction of gas pressure on the current in the leader stroke.

(*a*) 3 850-ohm series resistance.
(*b*) 1 850-ohm series resistance.

(*a*) 760mm Hg pressure.
(*b*) 593mm Hg pressure.
(*c*) 435mm Hg pressure.
(*d*) 227mm Hg pressure.

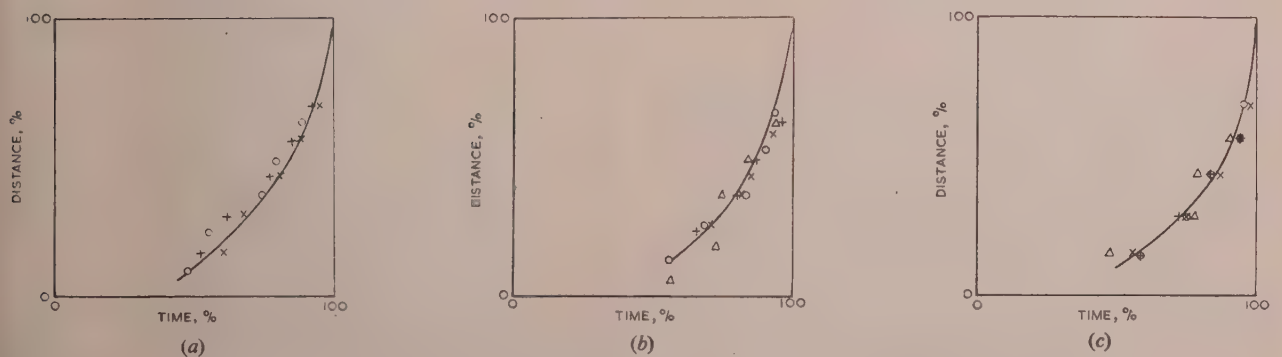


Fig. 13.—Effect of the reduction of gas pressure on the percentage-distance/percentage-time curve.

(*a*) 23-kilohm series resistance.
(*b*) 3 850-ohm series resistance.
(*c*) 1 850-ohm series resistance.

× 760mm Hg pressure.
\+ 593mm Hg pressure.
○ 435mm Hg pressure.
△ 227mm Hg pressure.

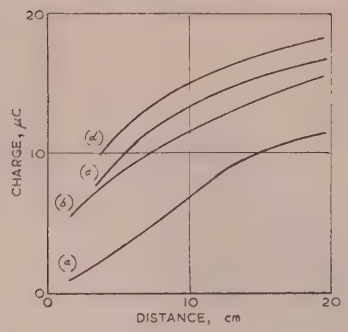


Fig. 14.—Effect of the reduction of gas pressure on the charge flow caused by the movement of the leader stroke.

(*a*) 760 mm Hg pressure.
(*b*) 593 mm Hg pressure.
(*c*) 435 mm Hg pressure.
(*d*) 227 mm Hg pressure.

It is obviously, however, greater than this, which suggests that the light emission from the leader stroke emanates from a diffuse source, and that the luminosity might also extend in a direction perpendicular to the leader stroke's direction of growth as well as in that direction. This view was confirmed by placing a slit parallel to the direction of growth of the leader stroke but to one side of its path. The curve so obtained of average peak light emission as a function of the distance of the slit from the spark path is shown in Fig. 15. This curve is necessarily rather

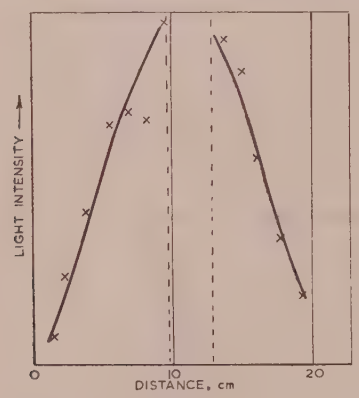


Fig. 15.—The distribution of light in the "head" of the leader stroke.

inaccurate, as individual leader strokes move along different paths. The distance shown as abscissa is therefore only an average distance for a reasonable number of records. The rapid rate of change of light with distance near to the bright central portion of the leader strokes makes it impossible with this technique to plot the central part of the curve shown in Fig. 15.

It has been stated that the percentage-distance/percentage-time curves exhibit a shape which is independent of the gap length if it is assumed that the initiation process commences at a finite time after the application of the voltage wave; it was therefore considered profitable to study the early stages of the initiation process. From Fig. 2 it will be seen that light output from the spark-gap increases very suddenly to a small value and then decreases to a negligible value, after which it rises again slowly as the bright part of the leader stroke approaches the slit. This initial light pulse (marked "corona" in Fig. 2) was found to occur in a gap subjected to a voltage insufficient to cause breakdown, and has been designated "impulse corona." Davis[26] had previously demonstrated that a current pulse of this general shape could be detected in the pointed electrode for a gap subjected to a voltage insufficient to cause breakdown. This impulse corona probably has a considerable bearing on the onset and initial growth of the leader stroke, and further measurements of its visual and electrical characteristics are described in the next Section.

(3.2) The Impulse Corona

An example of the light pulse caused by the impulse corona is shown in Fig. 16(*a*). It will be seen that the pulse has a wave-

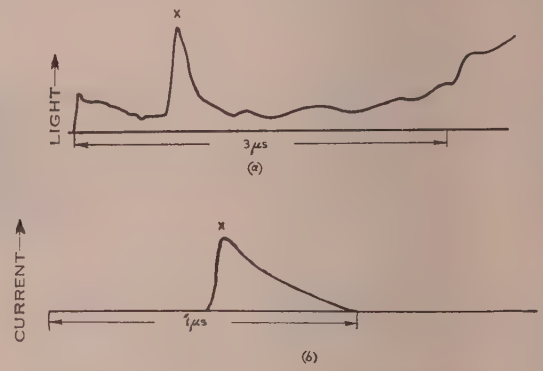


Fig. 16.—The light emission and current flow due to the corona pulse.

front of very short duration. The calibration of the photomultiplier-amplifier-oscillograph unit described above has shown that when a unit function is applied to this unit the output rises to approximately 95% of the full amplitude in approximately 6×10^{-8} sec. This is the time occupied by the recorded front of the corona pulse, and it is therefore considered safe to state that the corona pulse has a very steep front occupying a time not greater than 6×10^{-8} sec. Having reached this amplitude, the light then decays in a roughly exponential manner and falls to a small amplitude in approximately 4×10^{-7} sec. The duration of this pulse appears to be independent of its amplitude, of the voltage applied to the gap, of the circuit parameters, and of the gap length, to a first order of magnitude.

When a voltage of 50–60% of the minimum breakdown value is applied to a gap, the impulse corona consists of a single pulse which occurs some time after the voltage wave has reached its peak amplitude. As the voltage is increased, the corona pulse occurs earlier on the voltage waveform; also more than one pulse may occur for a single applied-voltage wave. At the minimum voltage necessary to cause a sparkover, several corona pulses may occur in rapid succession, with some overlapping. The last corona pulse in the sequence, instead of decaying completely, usually decays partly and then increases again. This increase in light emission, shown in Fig. 22 following the corona light pulse, corresponds to the light from the leader stroke and continues until the light from the main stroke causes the deflection to increase abruptly.

Similar effects are observed if the current in the circuit is recorded instead of the light emission. A typical oscillogram is shown in Fig. 16(*b*). The current pulse has the same general shape as the light pulse except that the front of the pulse occupies a slightly longer time, which may be due to the capacitive loading of the Metrosil.

The variation of the peak light emitted by the complete corona discharge as the amplitude of the voltage wave applied to the gap is varied is shown in Fig. 17. The series resistance in this case is 23 000 ohms. The full line in this Figure corresponds

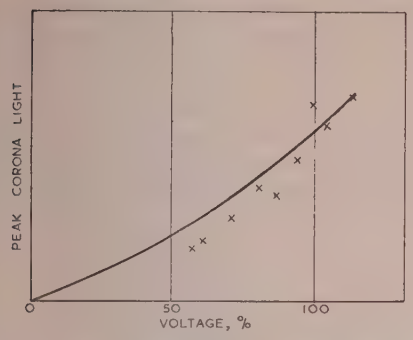


Fig. 17.—The effect of voltage on the peak light emitted by the corona pulse.
Voltage is shown as a percentage of the minimum breakdown voltage.

to the law $L = AV^{3/2}$, where L is the light emission and V is the applied voltage, the constant A being chosen arbitrarily. The variation of the peak current flowing in a corona pulse with

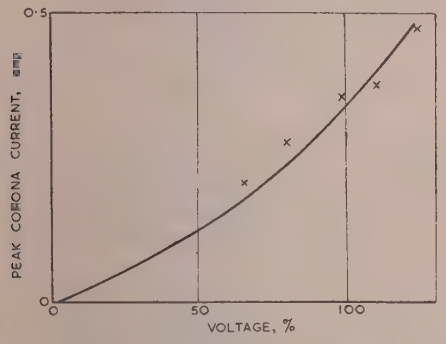


Fig. 18.—The effect of voltage on the peak current in the corona pulse.
Voltage is shown as a percentage of the minimum breakdown voltage.

change of applied voltage is shown in Fig. 18 and is seen to be similar to that of Fig. 17, with the full curve again corresponding to a law $I = BV^{3/2}$, in which I is the peak current.

Experiments were performed to determine the variation of the total corona light and current with variation of the series resistance. Over the range 10 000–40 000 ohms, no appreciable change of either light or current was recorded. Again it must be emphasized that changes sought here are first-order changes, and variations of about 5% or less are ignored. It is difficult to investigate the corona for low values of series resistance, as the voltage wavefront is then of very short duration. As a result, the corona pulse occurs shortly after the application of the voltage wave and tends to be confused with the stray pick-up which could not be eliminated. In the case of the corona light pulse, this tends to be confused with the short-duration cathode-ray-tube deflection caused by scattered light from the impulse-generator sparks.

It has been stated above that the leader stroke appears to consist of a large ill-defined volume of light emission. It was considered probable, therefore, that the impulse corona might exhibit the same sort of characteristics, and experiments were performed, similar to those for the leader stroke, to determine the physical characteristics of the impulse corona. It proved in some ways easier to deal with the impulse corona, since its position could be rather more readily defined. The curves showing the light emission through a slit as a function of the distance of the slit from the tip of the pointed electrode are

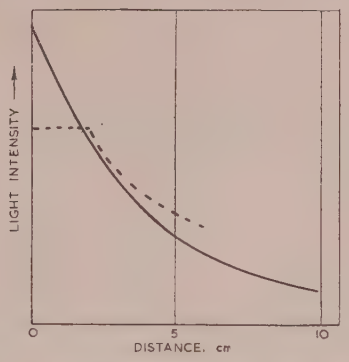


Fig. 19.—The distribution of light in the corona pulse.

shown in Fig. 19. The full curve corresponds to the light received through the slit perpendicular to the direction of the spark-gap, while the dotted curve corresponds to the light received through a slit parallel to the direction of the spark-gap. There is considerable similarity between these curves except for the dotted curve at zero distance. This difference may account for the fact that at this particular point the pointed electrode is probably obscuring part of the light.

The duration of the corona pulse, defined as the time which elapses between its inception and the decay of the light to about 10% of its peak value, as a function of gas pressure is shown

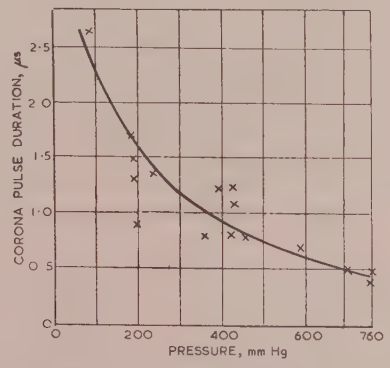


Fig. 20.—The effect of the reduction of gas pressure on the duration of the corona pulse.

in Fig. 20 for four different applied voltages. It will be seen that as the gas pressure is lowered the duration of the corona pulse increases.

(4) RESULTS IN GASES OTHER THAN AIR

The initiation mechanism was investigated for nitrogen, hydrogen and oxygen.

(4.1) Nitrogen

The form of the corona current pulse in nitrogen is shown in Fig. 21(*a*) and is seen to be of long duration compared with that in air. It is difficult to quote a figure for its duration, as will be obvious from an examination of Fig. 21(*a*). It can be stated, however, that the amplitude has fallen to a small value in a time of the order of 20 microsec.

The size of the corona and the distribution of light therein were not investigated.

An attempt was made to obtain the distance/time curve of the leader stroke in nitrogen. While a pulse of light could be

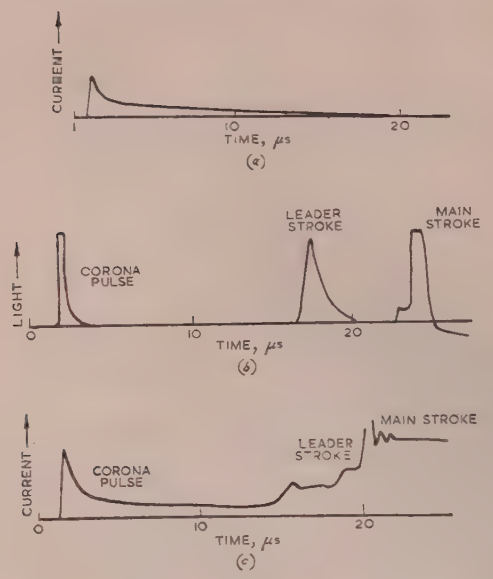


Fig. 21.—Corona pulse and leader stroke in nitrogen.

obtained as the leader stroke passed the slit, indicating that the leader stroke was growing in an orderly manner similar to that in air, the variations from leader stroke to leader stroke were so great that the measurements were felt to be meaningless. The mean time-lag for a 20 cm gap was approximately 19 microsec.

The form of the pulse of light when the leader stroke passes the slit is shown in Fig. 21(*b*) and is of a different shape from that obtained when studying the leader stroke in air. The inference from Fig. 21(*b*) is that the leader stroke is of much smaller dimensions in nitrogen.

Fig. 21(*c*) shows the current/time curve for the complete breakdown process, including the leader stroke and the main stroke.

(4.2) Hydrogen

The corona pulse for hydrogen, which is shown in Fig. 22(*a*), is similar to that in nitrogen.

In hydrogen, no evidence of leader strokes could be found. Fig. 22(*b*) shows the current in a gap subjected to a voltage causing breakdown as a function of time. The corona pulse forms early and has decayed to a small value when a sudden rise of current indicates the onset of the main stroke.

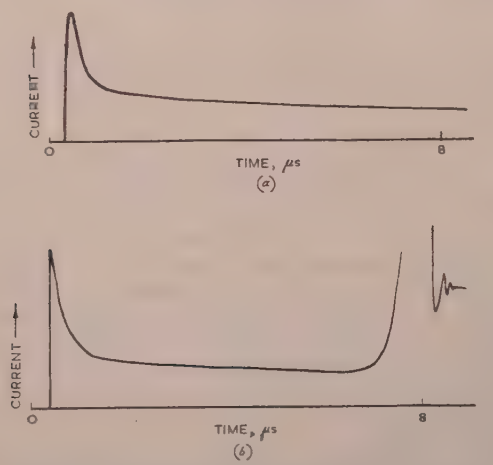


Fig. 22.—Corona pulse and breakdown in hydrogen.

(4.3) Oxygen

Fig. 23(*a*) shows the corona current pulse in oxygen. The pulse is of very short duration ($\sim 3 \times 10^{-8}$ sec) and is also of very small amplitude on the oscillogram. Since the resolution

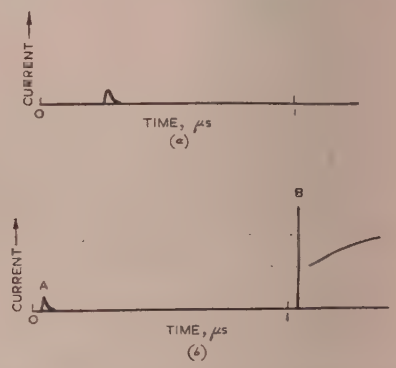


Fig. 23.—Corona pulse and breakdown in oxygen.

time of the recording equipment (6×10^{-8} sec) is longer than the recorded duration of this pulse ($\sim 3 \times 10^{-8}$ sec) it is certain that the time-pulse shape and amplitude were not recorded, and the most that can be claimed is that the existence of a short-duration ($> 6 \times 10^{-8}$ sec) small-amplitude pulse has been established.

Again no evidence could be found of leader strokes. Fig. 23(*b*) shows a typical breakdown, in which a small corona pulse occurs at A, and a sudden rise of current due to the main stroke occurs at B.

(5) DISCUSSION

From the experiments described above, a general view of the sequence of events in the breakdown mechanism may be obtained. The breakdown mechanism in a point-plane gap (positive point–negative plane) subjected to an impulse voltage appears to commence with the formation of impulse corona. This corona consists of a luminous zone of discharge which decreases in intensity with increasing distance from the point, but which is detectable up to a distance of approximately 10 cm from the point. If the voltage applied to the gap is sufficiently high, the discharge process advances from the point to the plane in the form of a leader stroke. The head—or most advanced portion—of the leader stroke consists of a luminous zone of discharge which is generally similar to that of the impulse corona. The bright central core of this zone traces out the main track of the leader stroke, as photographed in the experiments described in Reference 6, and the final spark channel seems to form along this track.

The experiments do not enable the structure of the luminous zone to be investigated, although there is evidence from experiments by other workers[9,10,18–22] to suggest that it consists of numerous filamentary streamers converging on to a bright central core. It is necessary, however, to emphasize that photographic evidence in this field must be interpreted with caution, as the recorded dimensions of an extended non-uniform source of light may depend markedly on the sensitivity of the system.

The current which is recorded on the oscillograph during the existence of the corona pulse may be considered to be due

entirely to the motion of electrons in the gap, as the velocities of the positive ions are much smaller than those of the electrons. The recorded current pulse has a total duration of 4×10^{-7} sec and a front of less than 5×10^{-8} sec. The true front may well be shorter than this, as the apparatus used was incapable of responding in a time less than 5×10^{-8} sec.

If this build-up of current is caused by the motion of electrons from the edge of the luminous zone to the pointed electrode, the average velocity of the electrons must be not less than 2×10^{8} cm/sec corresponding to an electron energy of 12 eV.

The spectrum of the corona light has been recorded and has been shown to consist almost entirely of the second positive bands of N_2, within the frequency limits imposed by the Ilford HP.3 emulsion. The excitation function for these bands has been determined by Langstroth,[25] who showed that the minimum electron energy necessary to produce any excitation is 13 eV. We may therefore conclude that even at considerable distances from the point a number of electrons attain the necessary 13 eV of energy. It would also appear possible that in 5×10^{-8} sec electrons may be able to move a considerable part of the distance from the edge to the centre of the luminous zone.

It is not clear how free electrons may be produced in regions so far from the pointed electrode that the normal field due to the applied voltage will be only about 1 kV/cm or less. There is, however, no evidence available to suggest that the corona grows from the outside inwards to the point rather than in the reverse direction.

If the corona is assumed to grow outwards from the centre, then two possibilities occur:

(*a*) The region immediately adjacent to the point, where the field is very high, may become ionized, and the resulting emission of photons, or other secondary processes, may cause the appearance of free electrons in the more-remote lower-field regions. However, it is difficult to imagine the production of such a short-duration front of the current pulse, since (i) the relaxation time of excited states or the recombination time for the production of the necessary photons would make this unlikely, and (ii) the current pulse would be expected to build up in the same way as the voltage on the gap due to the gradual extension of the high-field region.

(*b*) The corona may be triggered by some mechanism so far unknown, but which occurs in close proximity to the point and causes the simultaneous growth of a multitude of filamentary conducting channels away from the point. The field at the tip of each of these filaments will be much higher than that which would exist in the absence of the filaments. This might explain the zone's extension to regions of apparent low field strength. However, these filaments would need to grow over a distance of 10 cm in a time not greater than 5×10^{-8} sec, and no mechanism has yet been put forward to account for the simultaneous production of such a large number of filaments.

The extinction of the corona current when the voltage is just below the minimum value needed to cause breakdown is a phenomenon which needs explanation. Once the corona pulse starts to form there would appear to be a copious supply of electrons to promote further ionization, and whatever mechanism of propagation is assumed for the corona would presumably suffice for its maintenance. However, the current pulse rises to a peak and thereafter decays to zero. The time of decay appears to be inversely proportional to the amount of oxygen present (the nitrogen and hydrogen used were cylinder gases of industrial purity and probably contained approximately 0·4% oxygen), which suggests a decay mechanism associated with the formation of negative ions.

The close similarity between the corona pulse and the "head" of the leader stroke tends to indicate that the two are different aspects of the same phenomenon. The corona pulse forms round the pointed electrode at a voltage which is insufficient to cause further development leading to breakdown, but when the voltage is raised to a higher value this corona pulse propagates across the gap and is known as a leader stroke. When the leader stroke has crossed to the other electrode the full short-circuit current flows, the channel then becomes much more highly luminous, and breakdown is said to have occurred.

It will be noted from Figs. 5(*b*), 6, 17, 18, and 19 that the acceleration of the motion of the leader stroke across the gap follows a law which, for the minimum breakdown condition, is independent of the gap length, of the series resistance, and of the gas pressure within the ranges considered. If it is assumed that the leader stroke channel has a negligible impedance and acts effectively as an extension of the pointed electrode, its motion will be expected to be of the form observed, since the applied voltage is being impressed on an ever-decreasing gap as the leader stroke moves towards the plane. Essentially, therefore, the acceleration will be a characteristic function of the configuration of the system.

The velocity of the leader stroke at any point in its motion across the gap, if we now consider conditions other than the minimum breakdown voltage, is governed by the applied voltage, the series resistance, the initial gap length, and the distance which the leader stroke has travelled, assuming that the impedance of the leader stroke channel is negligible.

It appears from experiment that the voltage drop along the leader stroke channel is small. Fig. 24 shows distance/time

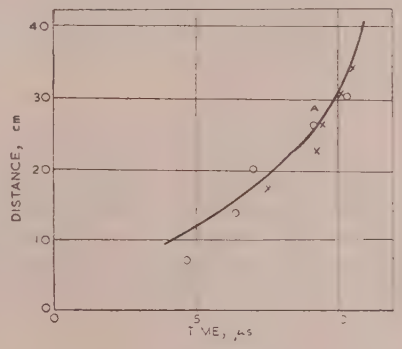


Fig. 24

——— 42·5 cm gap.
- - - 35 cm gap.
· · · 25 cm gap.

curves for gap lengths of 42·5, 35 and 25 cm, in each case with the minimum breakdown voltage for the 42·5 cm gap applied, plotted so as to coincide at the instant that the leader stroke reaches the plane electrode. It will be seen that the motion for most of the available gap is similar. For instance, at a distance of 10 cm from the plane electrode the leader stroke behaviour is the same for all three gap lengths. If there had been a voltage drop along the leader stroke channel equal to or greater than about 1 kV/cm, the conditions for the gap lengths of 35 cm and 25 cm when the leader stroke had reached the point marked A in Fig. 24 would have been different from those for the gap length of 42·5 cm, since the pointed electrode was in effect short-circuiting part of the leader stroke in the former cases. This is consistent with the measurements of Komelkov.

The effect of over-voltage on the gap may therefore be considered in the following way. If the gap length is l and the applied voltage is V, which is the minimum breakdown voltage of a gap length L, the leader stroke behaves as if it had moved a

distance ($L-l$) across a gap of length L subjected to a voltage V. Its motion from then on may therefore be predicted.

The effect of the series resistor arises from a lowering of the voltage on the gap due to the IR drop in it when a leader stroke starts to move and a current I flows in the gap. Since the current I is proportional to the velocity of the leader stroke, and since an increase in velocity causes an increase in I with a corresponding decrease in the applied voltage, it follows that at any point of the leader stroke's motion an equilibrium will tend to be set up. However, a phase shift will be introduced in the feedback loop due to the time-constant of the series resistor and the capacitance of the plane electrode, and this may cause the motion and the current to contain oscillatory components. In the limit, the oscillations may be large enough to cause the movement of the leader stroke to be periodically stopped—a phenomenon known as stepping.

It would appear that if the leader stroke is assumed to possess an inherent maximum velocity of propagation, this has not yet been measured, and that basically the motion of the leader stroke in the experiments described above has been controlled almost entirely by the circuit constants.

With the possible exception of the time during which the leader stroke is within a short distance of the plane electrode, it would appear that this electrode (the cathode) plays no part in the discharge mechanism other than the obvious role of the production of the required field. During the early stages of the discharge, the absence of light emission near the plane suggests that there is no ionization in this region. This is to be expected, moreover, since the field near the plane is quite small. Further, any electrons produced near the plane could be expected to move with such small velocities that it is improbable that they could influence the discharge mechanism near the anode in the short times involved.

(6) ACKNOWLEDGMENTS

The work was carried out with financial support from the British Electrical and Allied Industries Research Association, and the authors wish to thank Dr. S. Whitehead, Mr. R. Davis and other members of the relevant Sub-Committee for their continued interest and encouragement. They also acknowledge their appreciation of the helpful comments of Dr. J. D. Craggs during the progress of the work and in the preparation of the paper.

(7) REFERENCES

(1) Meek, J. M., and Craggs, J. D.: "Electrical Breakdown of Gases" (Oxford University Press, 1953).

(2) Schonland, B. F. J., and Collens, H.: "Progressive Lightning," *Proceedings of the Royal Society*, A, 1934, **143**, p. 654.

(3) Schonland, B. F. J., Malan, D. J., and Collens, H.: "Progressive Lightning II," *ibid.*, 1935, **152**, p. 595.

(4) Schonland, B. F. J.: "Progressive Lightning IV, The Discharge Mechanism," *ibid.*, 1938, **164**, p. 132.

(5) Schonland, B. F. J.: "The Pilot Streamer in Lightning and the Long Spark," *ibid.*, 1953, **220**, p. 25.

(6) Allibone, T. E., and Meek, J. M.: "The Development of the Spark Discharge," *ibid.*, 1938, **166**, p. 97, and 1938, **169**, p. 246.

(7) Komelkov, V. S.: "Testing of the Impulse Spark Flashover," *Elektrichestvo*, 1940, No. 9, pp. 24–27.

(8) Komelkov, V. S.: "An Investigation of Impulse Spark Discharges. I. The Question of Leader Velocities," *Journal of Technical Physics, U.S.S.R.*, 1940, **10**, p. 1426.

(9) Komelkov, V. S.: "The Leader Discharge Canal," *Dokl. Akad. Nauk., U.S.S.R.*, 1947, **68**, pp. 57–60.

(10) Komelkov, V. S.: "Structure and Parameters of the Leader Discharge," *Bulletin of the Academy of Science, U.S.S.R., Department of Technical Science*, No. 8, 1947, pp. 955–956.

(11) Komelkov, V. S.: "Development of the Electric Discharge in Long Gaps," *Isvestiya Akad. Nauk, U.S.S.R., Otdel teckh. Nauk*, No. 6, 1950, pp. 851–865.

(12) Komelkov, V. S., and Lifshits, A. M.: "Influence of Carriers on the Development of the Electric Discharge in Long Gaps," *ibid.*, No. 10, 1950, pp. 1463–1474.

(13) Wagner, C. F., McCann, G. D., and MacLane, G. L.: "Shielding of Transmission Lines," *Transactions of the American I.E.E.*, 1941, **60**, p. 313.

(14) Stekelnikov, I., and Beliakov, A.: "An Experimental Study of Spark Discharge," *Journal of Experimental and Theoretical Physics, Moscow*, 1940, **8**, No. 4, pp. 444–452.

(15) Stekolnikov, I.: "Investigations on Lightning in the Mountainous Districts of the Georgian S.S.R.," *Elektrichestvo*, 1940, No. 2, pp. 33–38.

(16) Stekolnikov, I.: "Measurement of the Propagation Rate of the Main Channel of a H.V. Discharge," *Dokl. Akad. Nauk*, 1952, **85**, No. 5, pp. 1013–1016.

(17) Akopian, A., and Larionov, V. P.: "Experimental Investigation of the Impulse Discharge in a Long Gap," *Elektrichestvo*, March, 1952, No. 3, pp. 31–33.

(18) Norinder, H., and Salka, O.: "Mechanism of Positive Spark Discharges with Long Gaps in Air at Atmospheric Pressure," *Arkiv för Fysik*, **3**, 1951, p. 347.

(19) Norinder, H., and Salka, O.: "Mechanism of Long-Gap Negative Spark Discharges in Air at Atmospheric Pressure," *ibid.*, **5**, 1952, p. 493.

(20) Norinder, H., and Salka, O.: "Screens in Long Discharge Gaps," *ibid.*, **6**, 1953, p. 151.

(21) Pucher, W.: "Negativer Durchschlag von Luft in der Funkenstrecke Kugel-Platte bei Stosspannungen," *ibid.*, p. 375.

(22) Pucher, W.: "Form der Negativen Vorentladung in Luft bei Atmosphärendruck für Stosspannungen," *ibid.*, p. 383.

(23) Saxe, R. F., and Meek, J. M.: "Development of Spark Discharges," *Nature*, 1948, **162**, p. 263.

(24) Prime, H. A., and Saxe, R. F.: "The Recording of Optical Transients," *Proceedings I.E.E.*, Paper No. 837 M, August, 1949 (**96**, Part II, p. 662).

(25) Langstroth, G. O.: "The Excitation of Band Systems by Electron Impact," *Proceedings of the Royal Society*, A, 1934, **146**, p. 166.

(26) Davis, R.: Private communication.

621.3.014.7

Monograph No. 125 S
Apr. 1955

MATRIX METHODS FOR THE EVALUATION OF SIMULTANEOUS FAULTS IN THREE-PHASE SYSTEMS

By W. E. LEWIS, Ph.D., and J. H. BANKS, B.Sc., Graduates.

(*The paper was first received 23rd July, 1954, and in revised form 29th November, 1954. It was published as an* INSTITUTION MONOGRAPH *in April, 1955.*)

SUMMARY

The paper illustrates two methods whereby the steady-state impedance matrix in terms of symmetrical component parameters may be obtained for a 3-phase system with multiple unsymmetrical faults. In the first method all faults are replaced by a balanced condition for which the symmetrical component impedance matrix is easily obtained using Stigant's rule. This matrix is then transformed to that of the faulted network by means of a connection matrix and relationships developed by Kron. The second method shows how Stigant's rule may be used to write down the impedance matrix of the symmetrical component network representing the unsymmetrical system.

LIST OF SYMBOLS

V, I, Z = Voltage, current and impedance matrices, respectively, of the system in terms of phase quantities.

V', I', Z' = Voltage, current and impedance matrices, respectively, in terms of symmetrical component parameters of the system with balanced faults.

V'', I'', Z'' = Voltage, current and impedance matrices, respectively, in terms of symmetrical component parameters of the system with unbalanced faults.

C = Connection matrix.

0, 1, 2 = Subscripts denoting zero, positive and negative phase sequence components respectively.

$a = -0{\cdot}5 + j0{\cdot}866$.

(1) INTRODUCTION

A 3-phase transmission system operating under balanced conditions (including symmetrical 3-phase faults) may be treated on a single-phase basis either analytically by the usual circuit theorems or experimentally by network analysers. When unsymmetrical faults occur, such as single line-to-earth, line-to-line, etc., the problem can be treated on a 3-phase basis, but since the impedance matrix of a 3-phase machine in terms of the line currents is not symmetrical, the circuit analysis is difficult and also a single-phase representation of the system is not possible.

The method usually adopted is to change the system of reference from phase quantities to the "symmetrical components" of Fortescue, for in this system the impedance matrices of machines and many of the faults encountered are symmetrical, thus permitting an equivalent single-phase circuit to be built up and solved in any convenient manner. In the analysis of single faults on an otherwise balanced system it is customary to define the fault so as to be symmetrical with respect to phase a, this being necessary if phase shifts are to be avoided in the interconnections of the sequence networks. In general, this simplifying device cannot be used for the analysis of multiple faults, for even if one fault is symmetrical with respect to phase a, some of the others may not be, and it is then necessary to interconnect the symmetrical-component sequence networks by a number of phase-shifting transformers. One method of analysis of such circuits has been given by Mortlock,[1] and is based on classical circuit theory.

The current tendency is towards a growing use of computing machines capable of performing any mathematical process which can be broken up into simple arithmetical operations. These machines are therefore eminently suitable for the addition, multiplication and inversion of matrices. Standard programmes of the operations to be performed by the machine have been derived, and include the inversion of matrices of very large order; these are immediately applicable to a particular problem, as it is only necessary to insert the actual matrix elements at the commencement of the solution. Descriptions of the construction, operation and programming of automatic digital computers have been given.[6,7,8] In order to use these machines for the solution of electrical circuits, the impedance and voltage matrices of the network are required. For many problems the impedance matrix is most easily obtained using Stigant's rule,[2] and Section 3.2.4 illustrates the application of the rule to the symmetrical-component equivalent circuit of a single line-to-earth fault on phase b in a system, which thus contains phase-shifting transformers in the interconnections.

In a large network with more than one unsymmetrical fault the delineation of the equivalent symmetrical-component circuit is not easy, and there is also the possibility of error in writing down the impedance and voltage matrices. Accordingly, a technique is put forward in Section 3.2.2 which enables these matrices to be derived from those pertaining to a related, but much simpler network, called the primitive network. The relations between the matrices of the primitive and actual networks have been deduced by Kron,[3] and are restated in Section 2.4. The primitive network used is the symmetrical-component equivalent circuit of the system, all unsymmetrical faults having been changed to a balanced condition. The zero-, positive- and negative-sequence networks are now separate and distinct from one another, and their impedance matrix is easily written down using Stigant's rule. In order to transform the matrices of the primitive network to those of the equivalent symmetrical-component circuit of the system with the unsymmetrical faults, a connection matrix is required which shows the relationships which exist between the sequence components of the fault currents by virtue of the unsymmetrical nature of the faults.

The engineers' task is now complete, for the solution of the problem follows routine matrix methods, and can be performed on a computer or, if the system is small, by a clerical staff.

(2) FUNDAMENTAL PROCESSES

(2.1) Stigant's Rule

Stigant's rule, which is virtually a modification of Kirchhoff's mesh law, is of great value in that it enables the impedance matrix of a network to be written down directly by inspection of

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Lewis is at the College of Technology, Birmingham.
Mr. Banks is with The General Electric Company, Limited, Witton.

the network. For the purposes of the paper its operation may be described as follows. The paths of the independent mesh currents are chosen to suit the problem under consideration, and are delineated on the network. The elements of the impedance matrix are then given by:

(*a*) A diagonal element such as Z_{rr} is the total impedance of the path of the rth mesh current.

(*b*) A non-diagonal element such as Z_{rn} is the impedance common to the rth and nth mesh currents. It is positive if I_r and I_n flow through the impedance in the same direction and negative if they flow in opposite directions. For linear bilateral networks, the impedance matrix is symmetrical and $Z_{rn} = Z_{nr}$.

As an example of the use of the rule, the equations of the network of Fig. 1 are written as

$$\begin{array}{|c|} \hline V_x \\ \hline V_x - V_y \\ \hline \end{array} = \begin{array}{|c|c|} \hline Z_x + Z_v & Z_x \\ \hline Z_x & Z_x + Z_y \\ \hline \end{array} \begin{array}{|c|} \hline I_x \\ \hline I_y \\ \hline \end{array} \quad . \quad (1)$$

directly from the network. It should be stressed that corresponding elements of the voltage and current matrices must refer to the same mesh.

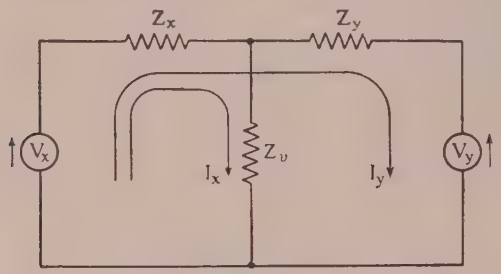


Fig. 1.—Two generators in parallel supplying a load.

(2.2) Symmetrical Components

In this Section, the zero-, positive- and negative-sequence currents and voltages are defined in terms of the corresponding phase quantities, and an expression is then deduced for the transformation of impedances from the phase-reference axes to the symmetrical-component reference axes. This transformation is then used to obtain the zero-, positive- and negative-sequence components of the phase impedances of the simple 3-phase systems which form the basis of the present study, namely machines, and balanced and unbalanced star-connected loads with impedance in the neutral. Reference 5 gives the usual approach to the problem.

(2.2.1) Definitions.

The symmetrical-component currents $I' = \{I_0, I_1, I_2\}$ and voltages $V' = \{V_0, V_1, V_2\}$ are defined in terms of the 3-phase currents $I = \{I_a, I_b, I_c\}$ and the line-to-neutral voltages $V = \{V_a, V_b, V_c\}$ by the relationships

$$I' = AI \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

and

$$V' = AV \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where

$$A = \tfrac{1}{3} \begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 1 & a & a^2 \\ \hline 1 & a^2 & a \\ \hline \end{array} \quad . \quad . \quad . \quad . \quad (4)$$

The inverse relationships are

$$I = A^{-1} I' \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

and

$$V = A^{-1} V' \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where

$$A^{-1} = \begin{array}{c|c|c|c|} & 0 & 1 & 2 \\ \hline a & 1 & 1 & 1 \\ \hline b & 1 & a^2 & a \\ \hline c & 1 & a & a^2 \\ \hline \end{array} \quad . \quad . \quad . \quad (7)$$

The transformation law of impedances follows. For if in the old (phase) system of reference

$$V = ZI$$

and in the new (symmetrical-component) system of reference

$$V' = Z'I'$$

then

$$V' = AV = AZI = AZA^{-1}I'$$

and thus

$$Z' = AZA^{-1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

It should be noted that using these definitions the apparent power, calculated from $V_t^* I$, is not the same in the two systems of reference. The transformation used by Kron differs from the one given here, in that the multiplier of the A matrix is $1/\sqrt{3}$ (instead of 1/3) and gives invariance of power.

(2.2.2) Symmetrical-Component Equivalent Circuits.

(*a*) *Machines.*—By making the usual assumptions of neglect of saturation, harmonics, etc., for power-system analysis, the impedance matrix of a 3-phase machine in terms of phase currents is circulant and of the form

$$Z' = \begin{array}{c|c|c|c|} & a & b & c \\ \hline a & Z_a & Z_b & Z_c \\ \hline b & Z_c & Z_a & Z_b \\ \hline c & Z_b & Z_c & Z_a \\ \hline \end{array} \quad . \quad . \quad . \quad (9)$$

which prevents the use of a single-phase equivalent circuit for the analysis of unbalanced conditions.

This impedance matrix may be transformed to the symmetrical-component matrix Z' by using the expression $Z' = AZA^{-1}$ [eqn. (8)], giving

$$Z' = \begin{array}{c|c|c|c|} & 0 & 1 & 2 \\ \hline 0 & Z_a + Z_b + Z_c & & \\ \hline 1 & & Z_a + a^2 Z_b + aZ_c & \\ \hline 2 & & & Z_a + aZ_b + a^2 Z_c \\ \hline \end{array} \quad . \quad . \quad . \quad . \quad (10)$$

We may write

$$Z_0 = Z_a + Z_b + Z_c$$
$$Z_1 = Z_a + a^2 Z_b + aZ_c$$
$$Z_2 = Z_a + aZ_b + a^2 Z_c$$

This matrix has elements on the main diagonal only, and consequently the equivalent circuit shown in Fig. 2 consists of

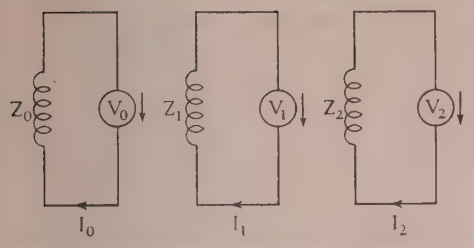


Fig. 2.—Symmetrical-component equivalent circuit of a 3-phase machine.

three separate networks, so that the mutual coupling is eliminated; the calculation of the currents due to unbalanced applied voltages may therefore be readily carried out.

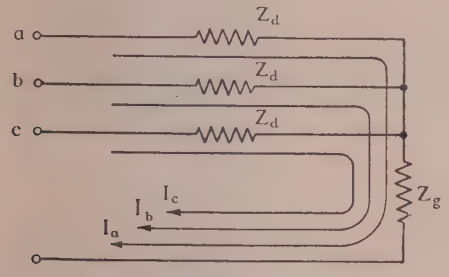


Fig. 3.—Balanced 3-phase load.

(*b*) *Balanced Load.*—Fig. 3 shows a balanced load of Z_d ohms per phase and a neutral impedance of Z_g ohms. In terms of the phase currents the impedance matrix is

$$\boldsymbol{Z} = \begin{array}{c|ccc} & a & b & c \\ \hline a & Z_d + Z_g & Z_g & Z_g \\ b & Z_g & Z_d + Z_g & Z_g \\ c & Z_g & Z_g & Z_d + Z_g \end{array} \quad . \; . \; . \quad (11)$$

The transformation to the symmetrical-component reference system gives $\boldsymbol{Z}' = \boldsymbol{AZA}^{-1}$, or

$$\boldsymbol{Z}' = \begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & Z_d + 3Z_g & & \\ 1 & & Z_d & \\ 2 & & & Z_d \end{array} \quad . \; . \; . \quad (12)$$

Once more the mutual coupling is removed. The equivalent circuit is shown in Fig. 4.

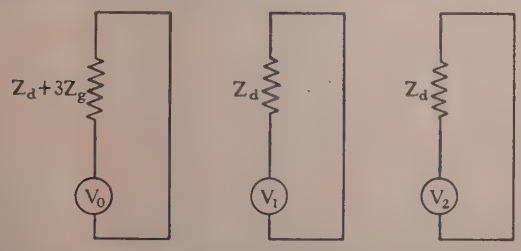


Fig. 4.—Symmetrical-component equivalent circuit of the balanced 3-phase load of Fig. 3.

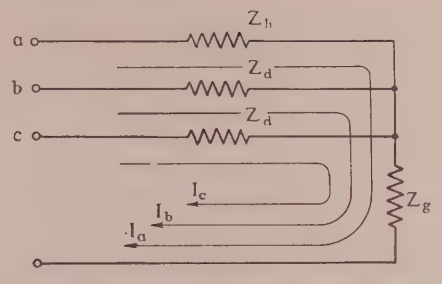


Fig. 5.—Partially unbalanced 3-phase load.

(*c*) *Partially Unbalanced Load.*—Fig. 5 shows a partially unbalanced star-connected load in which the impedance in two of the phases is Z_d and in the third phase Z_h. An impedance Z_g is included in the neutral lead. In terms of the phase currents shown, the impedance matrix is

$$\boldsymbol{Z} = \begin{array}{c|ccc} & a & b & c \\ \hline a & Z_h + Z_g & Z_g & Z_g \\ b & Z_g & Z_d + Z_g & Z_g \\ c & Z_g & Z_g & Z_d + Z_g \end{array} \quad . \; . \; . \quad (13)$$

and using the transformation $\boldsymbol{Z}' = \boldsymbol{AZA}^{-1}$ the symmetrical-component impedance matrix is

$$\boldsymbol{Z}' = \frac{1}{3} \begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & Z_h + 2Z_d + 9Z_g & Z_h - Z_d & Z_h - Z_d \\ 1 & Z_h - Z_d & Z_h + 2Z_d & Z_h - Z_d \\ 2 & Z_h - Z_d & Z_h - Z_d & Z_h + 2Z_d \end{array} \quad . \; . \; . \quad (14)$$

The equivalent circuit of this matrix is shown in Fig. 6. If the applied voltages are balanced, V_0 and V_2 are both zero and the currents are found easily.

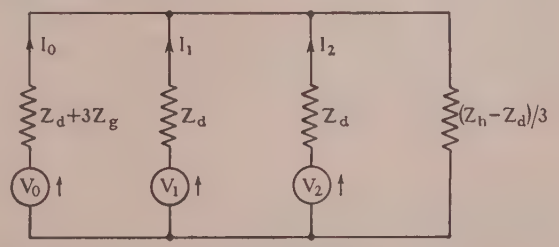


Fig. 6.—Symmetrical-component equivalent circuit of the load shown in Fig. 5.

(*d*) *Completely Unbalanced Load.*—Fig. 7 shows an unbalanced star load having impedances in each of the three phases of Z_a, Z_b and Z_c ohms, respectively. An impedance of Z_g is included in the neutral lead.

The impedance matrix in terms of the three line currents is

$$\boldsymbol{Z} = \begin{array}{c|ccc} & a & b & c \\ \hline a & Z_a + Z_g & Z_g & Z_g \\ b & Z_g & Z_b + Z_g & Z_g \\ c & Z_g & Z_g & Z_c + Z_g \end{array} \quad . \; . \; . \quad (15)$$

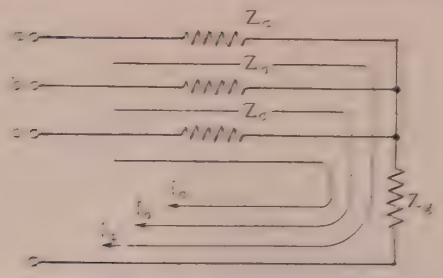


Fig. 7.—Completely unbalanced 3-phase load.

Transforming this matrix to symmetrical components by $Z' = AZA^{-1}$ gives,

$$Z' = \begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & Z_1 - Z_2 - Z_c + 9Z_g & Z_1 - a^2Z_2 - aZ_c & Z_a + aZ_b - a^2Z_c \\ 1 & Z_1 - aZ_2 - a^2Z & Z_a - Z_b - Z & Z_a + a^2Z_b - aZ_c \\ 2 & Z_1 - a^2Z_2 - aZ_c & Z_a + aZ_b - a^2Z_c & Z_a - Z_b - Z_c \end{array} \quad \ldots\ldots (16)$$

In this case the transformation to the symmetrical-component reference system has complicated the matrix and made it unsymmetrical. Further details of the matrix treatment of symmetrical components will be found in Reference 4.

(2.3) Generated Voltages

The usual assumption is made in this paper, that positive-sequence voltages only are generated. However, the further common assumption—that all generators have equal generated voltages—is not used.

(2.4) Interconnection of Networks

Kron has shown that if the equations of a network ($V' = Z'I'$) are known, the equations of another network, derived from the first (primitive) by changing the connections, but without increasing the number of independent meshes, may be obtained in the following manner. The relationship between the mesh currents of the primitive network, I', and the mesh currents of the new network, I'', is written as

$$I' = CI'' \quad \ldots\ldots (17)$$

C being called the "connection matrix."

The new voltage matrix V'' is given by

$$V'' = \overline{C}_t V' \quad \ldots\ldots (18)$$

and the new impedance matrix Z'' is given by

$$Z'' = \overline{C}_t Z' C \quad \ldots\ldots (19)$$

where $\overline{C}_t$ is the complex conjugate of the transpose of C.

(3) EVALUATION OF FAULT CURRENTS

(3.1) Symmetrical Three-Phase Fault

No difficulty is encountered in evaluating this type of fault, for the system reduces to a single-phase network which can be solved on a network analyser, or the matrix equation can be written out for solution on a computer.

(3.2) Methods with Unsymmetrical Faults

(3.2.1) Classical Method.

In this method, the relationships between the line currents and line voltages are ascertained from the 3-phase circuit. These relationships are then expressed in terms of symmetrical components, and the required sequence network is deduced and solved in any convenient manner. When more than one unsymmetrical fault exists on the system, it is necessary to interconnect the sequence network, either by 1 : 1 ratio transformers, or by phase-shifting transformers, and the circuit becomes very complicated.

One method of solving such networks has been given by Mortlock,[1] and an alternative method using matrix techniques is given in Section 3.4.

(3.2.2) General Constraint Technique.

The zero-, positive- and negative-sequence networks for a 3-phase system with balanced-phase impedances are separate and distinct from one another, and the positive-sequence network is the single-phase representation of the system. Using Stigant's rule it is comparatively easy to write down the impedance matrix for such a network. Thus the primitive network chosen for the calculation of the steady-state fault currents is the symmetrical-component representation of the system, all faults having been replaced by a balanced condition. The impedance matrix of the primitive network will be denoted by Z', and the voltage and current matrices by V' and I' respectively. A connection matrix C is obtained by expressing in the form $I' = CI''$ the relationships between the mesh currents of the primitive network I' and the mesh currents actually necessary for the analysis I''. The number of currents in I'' will be less than those in I' because of the relationships existing between the sequence currents pertaining to an unsymmetrical fault.

Eqns. (18) and (19) enable the voltage and impedance matrices of the symmetrical-component representation of the unbalanced system to be calculated. The required currents may now be evaluated, either by hand using a desk calculating machine for small problems, or by an automatic digital computer for larger problems.

As a simple illustration of the method, consider a generator having line a short-circuited to neutral as in Fig. 8. In this

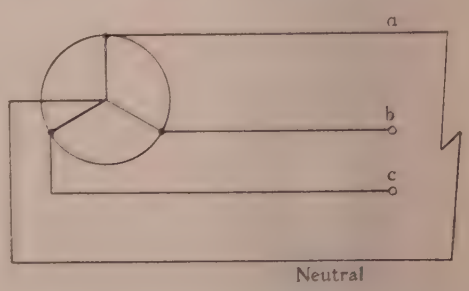


Fig. 8.—Three-phase generator with single line-to-neutral short-circuit.

technique the unsymmetrical fault is replaced by a suitable balanced condition, namely the symmetrical short-circuit to neutral shown in Fig. 9.

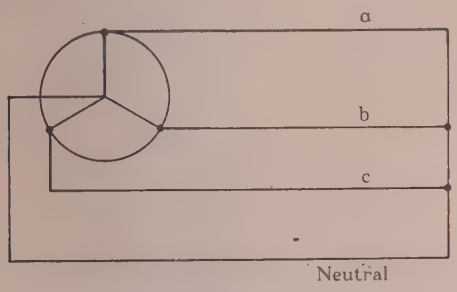


Fig. 9.—Three-phase generator with balanced short-circuit.

Sections 2.2.2 (*a*) and (*b*) show that the symmetrical-component network equivalent to Fig. 9 will be that shown in Fig. 10. The matrix equation of this network $V' = Z'I'$ will be

$$\begin{bmatrix} 0 \\ V_1 \\ 0 \end{bmatrix} = \begin{bmatrix} Z_0 & & \\ & Z_1 & \\ & & Z_2 \end{bmatrix} \begin{bmatrix} I_0 \\ I_1 \\ I_2 \end{bmatrix} \quad . \quad (20)$$

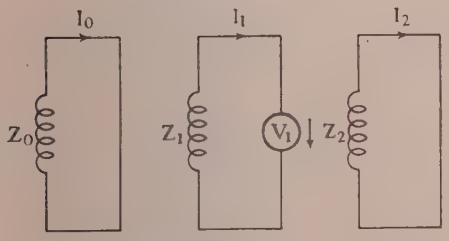


Fig. 10.—Symmetrical-component equivalent circuit of 3-phase generator with balanced short-circuit.

The connection matrix is developed in the following manner. In the original system of Fig. 9

$$I_b = I_c = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

Consequently from the definitions of the currents I_0, I_1, I_2, it is deduced that

$$I_0 = I_1 = I_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

and the analysis may proceed in terms of one current only, for example I_1.

The relationships are then written as

$$\begin{bmatrix} I_0 \\ I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} I_1 \end{bmatrix} \quad . \quad . \quad . \quad (23)$$

which is the equation

$$I' = CI'' \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

so that the connection matrix C is given by

$$C = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

The equations of the unsymmetrical fault are now determined from eqns. (18) and (19), and are,

$$V'' = \overline{C}_t V' = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ V_1 \\ 0 \end{bmatrix}$$

Therefore

$$V'' = V_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

and

$$Z'' = \overline{C}_t Z' C = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} Z_0 & & \\ & Z_1 & \\ & & Z_2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Therefore

$$Z'' = Z_0 + Z_1 + Z_2 \quad . \quad . \quad . \quad . \quad . \quad (27)$$

Also

$$I'' = I_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

In this simple case, there is one equation only, namely

$$V_1 = (Z_0 + Z_1 + Z_2)I_1 \quad . \quad . \quad . \quad . \quad (29)$$

from which I_1 may be calculated.

The equivalent circuit is seen to be that of Fig. 11.

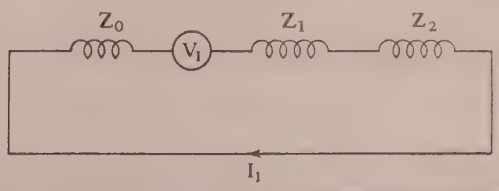


Fig. 11.—Symmetrical-component equivalent circuit of 3-phase generator with single line-to-neutral short-circuit.

(3.3) Applications involving Unsymmetrical Faults

In this Section the method will be applied to a number of typical single faults to show how the required balanced condition is decided, and also how the connection matrix is deduced. The system used for this explanation is shown in Fig. 12, and

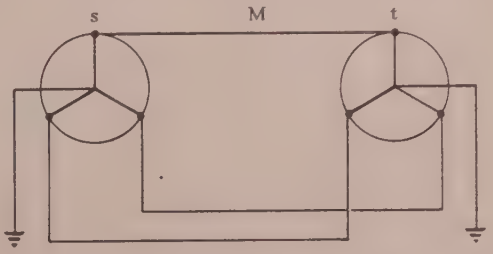


Fig. 12.—Simple system consisting of two generators connected by a tie line used to illustrate the constraint technique. Faults will be imposed at M.

consists of two generators both of which are solidly earthed at their neutrals, and connected by a tie line. The faults will be imposed at M.

(3.3.1) **Single Line-to-Earth Fault.**

Consider a single line-to-earth fault at M in Fig. 12, on line *a* having an arc impedance Z_f and a total ground impedance Z_g. This fault is replaced by a suitable balanced condition, which in this case is the load of Section 2.2.2(*b*), Fig. 3, in which $Z_d = Z_f$. The sequence networks for this balanced condition are shown

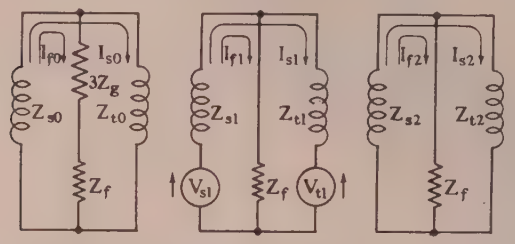


Fig. 13.—Symmetrical-component equivalent circuit of the system of Fig. 12, having a balanced fault at M. The arc impedance is Z_f ohms and the impedance in the neutral is Z_g ohms.

in Fig. 13, and using Stigant's rule the matrix equation of the network in terms of the mesh currents shown is

$$\begin{array}{|c|} \hline 0 \\ \hline 0 \\ \hline V_{s1} \\ \hline V_{s1} - V_{t1} \\ \hline 0 \\ \hline 0 \\ \hline \end{array} = \begin{array}{|c|c|c|c|c|c|} \hline Z_{s0} + Z_f + 3Z_g & Z_{s0} & & & & \\ \hline Z_{s0} & Z_{s0} + Z_{t0} & & & & \\ \hline & & Z_{s1} + Z_f & Z_{s1} & & \\ \hline & & Z_{s1} & Z_{s1} + Z_{t1} & & \\ \hline & & & & Z_{s2} + Z_f & Z_{s2} \\ \hline & & & & Z_{s2} & Z_{s2} + Z_{t2} \\ \hline \end{array} \begin{array}{|c|} \hline I_{f0} \\ \hline I_{s0} \\ \hline I_{f1} \\ \hline I_{s1} \\ \hline I_{f2} \\ \hline I_{s2} \\ \hline \end{array} \quad . \quad . \quad (30)$$

i.e. $\boldsymbol{V}' = \boldsymbol{Z}'\boldsymbol{I}'$.

Because of the unsymmetrical nature of the fault, namely line-*a*-to-ground, we have

$$I_{f0} = I_{f1} = I_{f2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

No such relationship exists between the other three currents, since the meshes in which they flow are unaffected by the fault. These facts are then expressed as in eqn. (32), the analysis proceeding in terms of I_{f0}, I_{s0}, I_{s1} and I_{s2},

$$\begin{array}{|c|} \hline I_{f0} \\ \hline I_{s0} \\ \hline I_{f1} \\ \hline I_{s1} \\ \hline I_{f2} \\ \hline I_{s2} \\ \hline \end{array} = \begin{array}{|c|c|c|c|} \hline 1 & & & \\ \hline & 1 & & \\ \hline 1 & & & \\ \hline & & 1 & \\ \hline 1 & & & \\ \hline & & & 1 \\ \hline \end{array} \begin{array}{|c|} \hline I_{f0} \\ \hline I_{s0} \\ \hline I_{s1} \\ \hline I_{s2} \\ \hline \end{array} \quad (32)$$

i.e. $\boldsymbol{I}' = \boldsymbol{C}\boldsymbol{I}'' \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$

$\boldsymbol{C}$ is thereby defined for this problem.

The required voltage and impedance matrices are then obtained in the usual manner, namely

$$\boldsymbol{V}'' = \overline{\boldsymbol{C}}_t\boldsymbol{V}' = \begin{array}{c|c} & \\ f_0 & V_{s1} \\ \hline s_0 & 0 \\ \hline s_1 & V_{s1} - V_{t1} \\ \hline s_2 & 0 \end{array} \quad . \quad . \quad . \quad (34)$$

$$\boldsymbol{Z}'' = \overline{\boldsymbol{C}}_t\boldsymbol{Z}'\boldsymbol{C}$$

$$\boldsymbol{Z}'' = \begin{array}{c|c|c|c|c} & f_0 & s_0 & s_1 & s_2 \\ \hline f_0 & Z_{s0} + Z_{s1} + Z_{s2} + 3(Z_g + Z_f) & Z_{s0} & Z_{s1} & Z_{s2} \\ \hline s_0 & Z_{s0} & Z_{s0} + Z_{t0} & & \\ \hline s_1 & Z_{s1} & & Z_{s1} + Z_{t1} & \\ \hline s_2 & Z_{s2} & & & Z_{s2} + Z_{t2} \\ \hline \end{array} \quad . \quad . \quad . \quad . \quad (35)$$

$\boldsymbol{I}''$ is defined by eqns. (32) and (33).

The equivalent network of the equation $\boldsymbol{V}'' = \boldsymbol{Z}''\boldsymbol{I}''$ is shown in Fig. 14.

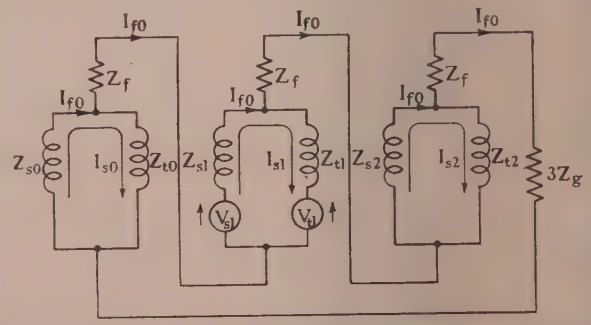


Fig. 14.—Symmetrical-component equivalent circuit of the system of Fig. 12, having a fault of total impedance $Z_f + Z_g$ ohms, between line *a* and earth at M.

If the fault was on line *b* or line *c*, and it was the only fault, the analysis could proceed as above, but if this was one of a number of faults, it would be essential to consider the faults on the lines on which they occurred.

For example, suppose the fault was on line b. The relationship between the sequence currents in the fault mesh is

$$I_{f1} = a^2 I_{f2} = a I_{f0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

(from $I_{fa} = 0$ and $I_{fc} = 0$).

Thus the connection matrix in terms of I_{f0} and I_{s0}, I_{s1} and I_{s2} is

$$C = \begin{array}{c|cccc} & f_0 & s_0 & s_1 & s_2 \\ \hline f_0 & 1 & & & \\ s_0 & & 1 & & \\ f_1 & a & & & \\ s_1 & & & 1 & \\ f_2 & a^2 & & & \\ s_2 & & & & 1 \end{array} \quad . \quad . \quad (37)$$

$$\overline{C}_t = \begin{array}{c|cccccc} & f_0 & s_0 & f_1 & s_1 & f_2 & s_2 \\ \hline f_0 & 1 & & a^2 & & a & \\ s_0 & & 1 & & & & \\ s_1 & & & & 1 & & \\ s_2 & & & & & & 1 \end{array}$$

The voltage, current, and impedance matrices for the balanced condition are unaltered, consequently

$$\overline{V}'' = [C_t]V' = \begin{array}{c|c} & \\ f_0 & a^2 V_{s1} \\ s_0 & 0 \\ s_1 & V_{s1} - V_{t1} \\ s_2 & 0 \end{array} \quad . \quad . \quad . \quad . \quad . \quad (38)$$

$$Z'' = \overline{C}_t Z' C$$

$$= \begin{array}{c|cccc} & f_0 & s_0 & s_1 & s_2 \\ \hline f_0 & Z_{s0} + Z_{s1} + Z_{s2} + 3(Z_g + Z_f) & Z_{s0} & a^2 Z_{s1} & a Z_{s2} \\ s_0 & Z_{s0} & Z_{s0} + Z_{t0} & & \\ s_1 & a Z_{s1} & & Z_{s1} + Z_{t1} & \\ s_2 & a^2 Z_{s2} & & & Z_{s2} + Z_{t2} \end{array} \quad . \quad . \quad . \quad . \quad (39)$$

It will be seen that this matrix is unsymmetrical. Its equivalent circuit is shown in Fig. 22.

The above example shows the ease with which the equations for the unsymmetrical fault may be obtained, since neither the impedance matrix for the symmetrical condition nor the connection matrix presents any difficulty.

(3.3.2) **Double Line-to-Earth Fault.**

Consider a double line-to-earth fault at M on the system of Fig. 12, with a total impedance in the earth path of Z_g and negligible impedance between the lines and earth. The balanced

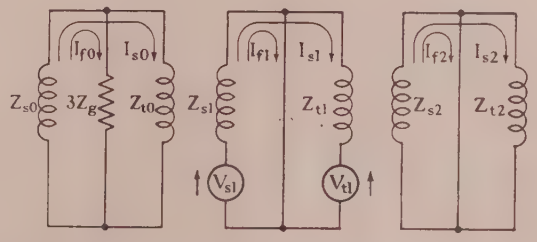


Fig. 15.—Symmetrical-component equivalent circuit of the system of Fig. 12, having a balanced fault of negligible arc impedance and Z_g ohms in the earth path imposed at M.

condition used is once more that of Fig. 3, but with $Z_d = 0$. The equivalent circuit of the system in terms of symmetrical components is shown in Fig. 15. Its impedance matrix is

$$Z' = \begin{array}{c|cccccc} & f_0 & s_0 & f_1 & s_1 & f_2 & s_2 \\ \hline f_0 & Z_{s0} + 3Z_g & Z_{s0} & & & & \\ s_0 & Z_{s0} & Z_{s0} + Z_{t0} & & & & \\ f_1 & & & Z_{s1} & Z_{s1} & & \\ s_1 & & & Z_{s1} & Z_{s1} + Z_{t1} & & \\ f_2 & & & & & Z_{s2} & Z_{s2} \\ s_2 & & & & & Z_{s2} & Z_{s2} + Z_{t2} \end{array} \quad . \quad . \quad (40)$$

and the voltage matrix is

$$V' = \begin{array}{c|c} f_0 & 0 \\ s_0 & 0 \\ f_1 & V_{s1} \\ s_1 & V_{s1} - V_{t1} \\ f_2 & 0 \\ s_2 & 0 \end{array} \quad \ldots\ldots \quad (41)$$

The connection matrix will depend on which two lines are faulted, but in any case, one of the sequence components of the fault mesh current can be expressed in terms of the other two.

(*a*) *Fault on Lines b and c.*—From the definitions [eqns. (2) and (4)] we see that, if the fault is on lines *b* and *c*, the relationships between the currents in the fault mesh is

$$I_{f2} = -(I_{f1} + I_{f0}) \quad \ldots\ldots \quad (42)$$

There is no relationship between the currents in the *s* mesh.

The connection matrix C is then given by

$$C = \begin{array}{c|ccccc} & f_0 & s_0 & f_1 & s_1 & s_2 \\ \hline f_0 & 1 & & & & \\ s_0 & & 1 & & & \\ f_1 & & & 1 & & \\ s_1 & & & & 1 & \\ f_2 & -1 & & -1 & & \\ s_2 & & & & & 1 \end{array} \quad \ldots \quad (43)$$

The required impedance matrix $Z'' = \bar{C}_t Z' C$ is then

$$Z'' = \begin{array}{c|ccccc} & f_0 & s_0 & f_1 & s_1 & s_2 \\ \hline f_0 & Z_{s0} + Z_{s2} + 3Z_g & Z_{s0} & Z_{s2} & & -Z_{s2} \\ s_0 & Z_{s0} & Z_{s0} + Z_{t0} & & & \\ f_1 & Z_{s2} & & Z_{s1} + Z_{s2} & Z_{s1} & -Z_{s2} \\ s_1 & & & Z_{s1} & Z_{s1} + Z_{t1} & \\ s_2 & -Z_{s2} & & -Z_{s2} & & Z_{s2} + Z_{s2} \end{array} \quad \ldots\ldots \quad (44)$$

and the voltage matrix $V'' = [C_t]V'$ is

$$V'' = \begin{array}{c|c} f_0 & 0 \\ s_0 & 0 \\ f_1 & V_{s1} \\ s_1 & V_{s1} - V_{t1} \\ s_2 & 0 \end{array} \quad \ldots\ldots \quad (45)$$

(*b*) *Fault on Lines a and b.*—The equation of current constraint in the fault mesh is now

$$I_0 + aI_1 + a^2 I_2 = 0 \quad \ldots\ldots \quad (46)$$

so that

$$I_{f2} = -(a^2 I_{f1} + aI_{f0}) \quad \ldots\ldots \quad (47)$$

and the connection matrix C will now be given by

$$C = \begin{array}{c|ccccc} & f_0 & s_0 & f_1 & s_1 & s_2 \\ \hline f_0 & 1 & & & & \\ s_0 & & 1 & & & \\ f_1 & & & 1 & & \\ s_1 & & & & 1 & \\ f_2 & -a & & -a^2 & & \\ s_2 & & & & & 1 \end{array} \quad \ldots \quad (48)$$

The required impedance matrix $Z'' = \bar{C}_t Z' C$ is

$$Z'' = \begin{array}{c|ccccc} & f_0 & s_0 & f_1 & s_1 & s_2 \\ \hline f_0 & Z_{s0} + Z_{s2} + 3Z_g & Z_{s0} & aZ_{s2} & & -a^2 Z_{s2} \\ s_0 & Z_{s0} & Z_{s0} + Z_{t0} & & & \\ f_1 & a^2 Z_{s2} & & Z_{s1} + Z_{s2} & Z_{s1} & -aZ_{s2} \\ s_1 & & & Z_{s1} & Z_{s1} + Z_{t1} & \\ s_2 & -aZ_{s2} & & -a^2 Z_{s2} & & Z_{s2} + Z_{t2} \end{array} \quad \ldots \quad (49)$$

and the voltage matrix is again

$$V'' = \begin{array}{c|c} f_0 & 0 \\ s_0 & 0 \\ f_1 & V_{s1} \\ s_1 & V_{s1} - V_{t1} \\ s_2 & 0 \end{array} \quad \ldots\ldots \quad (50)$$

(3.3.3) **Open Conductor.**

Consider the system of Fig. 12, with an open conductor on phase *a*, as shown in Fig. 16. There are two possible symmetrical conditions from which this fault may be derived.

(*a*) Machines disconnected and each one symmetrically short-circuited.

(*b*) Machines completed connected.

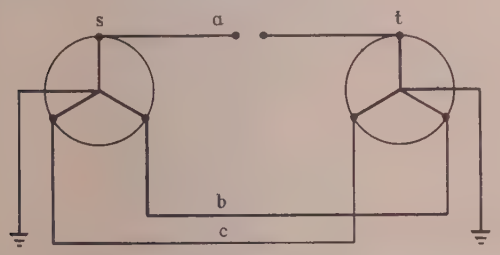


Fig. 16.—The system of Fig. 12 having an open conductor on line *a*.

(*a*) *Machines Disconnected and Short-Circuited.*—From Section 2.2.2(*a*) the impedance matrix for this system in terms of symmetrical components is

$$Z' = \begin{array}{cc|cccccc} & \textit{Generator} & & s & & & t & \\ & \textit{Sequence} & 0 & 1 & 2 & 0 & 1 & 2 \\ \hline s & 0 & Z_{s0} & & & & & \\ & 1 & & Z_{s1} & & & & \\ & 2 & & & Z_{s2} & & & \\ t & 0 & & & & Z_{t0} & & \\ & 1 & & & & & Z_{t1} & \\ & 2 & & & & & & Z_{t2} \end{array} \quad (51)$$

The connection matrix is deduced from the fact that $I_a = 0$, or in terms of symmetrical components $I_0 + I_1 + I_2 = 0$. Thus we may put

$$I_2 = -(I_0 + I_1) \quad \ldots\ldots \quad (52)$$

This constraint equation is the same as that for a double line-to-earth fault, but the circuit to which it is applied is different. The analysis of the fault will be taken in terms of the symmetrical component currents I_{s0} and I_{s1} of generator *s*. Then, since the assumed positive direction of current is out of a generator, the connection matrix is

$$C = \begin{array}{cc|cc} \textit{Generator} & & s & \\ & \textit{Sequence} & 0 & 1 \\ \hline s & 0 & 1 & \\ & 1 & & 1 \\ & 2 & -1 & -1 \\ t & 0 & -1 & \\ & 1 & & -1 \\ & 2 & 1 & 1 \end{array} \quad \ldots\ldots \quad (53)$$

The required impedance matrix $Z'' = \bar{C}_t Z' C$ is then

$$Z'' = \begin{array}{c|cc} & s_0 & s_1 \\ \hline s_0 & Z_{s0} + Z_{s2} + Z_{t0} + Z_{t2} & Z_{s2} + Z_{t2} \\ s_1 & Z_{s2} + Z_{t2} & Z_{s1} + Z_{s2} + Z_{t1} + Z_{t2} \end{array} \quad (54)$$

(*b*) *Both Machines Connected.*—The sequence networks for this symmetrical system are shown in Fig. 17.

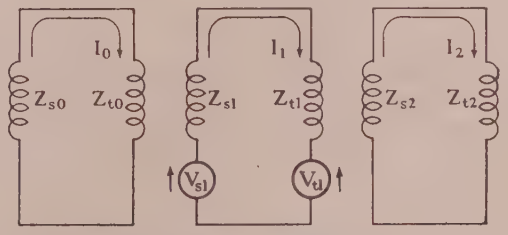


Fig. 17.—Symmetrical-component equivalent circuit of Fig. 12, no fault being imposed.

In terms of the currents shown, the impedance matrix is

$$Z' = \begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & Z_{s0} + Z_{t0} & & \\ 1 & & Z_{s1} + Z_{t1} & \\ 2 & & & Z_{s2} + Z_{t2} \end{array} \quad \ldots\ldots \quad (55)$$

The constraint equation is the same as before, namely $I_0 + I_1 + I_2 = 0$, so that analysing in terms of the currents I_0 and I_1, the connection matrix C is given by

$$C = \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1 & \\ 1 & & 1 \\ 2 & -1 & -1 \end{array} \quad \ldots\ldots \quad (56)$$

The impedance matrix for the fault condition is then cal-

culated in the usual way, and is found to be that given above, namely

$$\boldsymbol{Z''} = \begin{array}{c|cc} & s_0 & s_1 \\ \hline s_0 & Z_{s0}+Z_{t0}+Z_{s2}+Z_{t2} & Z_{s2}+Z_{t2} \\ s_1 & Z_{s2}+Z_{t2} & Z_{s1}+Z_{t1}+Z_{s2}+Z_{t2} \end{array} \quad (57)$$

(3.3.4) **Open Conductor and Line-to-Earth Fault.**

This fault occurs when one end of the broken conductor comes down to earth. Consider the same system as before, with such a fault at M on line a as shown in Fig. 18, the impedance of the fault path being composed of the arc impedance Z_f and an earth impedance Z_g.

The balanced condition chosen is that of both machines connected with a balanced star load of phase impedance Z_f and neutral impedance Z_g at the fault point M. This is the same condition as that used in Section 3.3.1, so that the corresponding sequence network is shown in Fig. 13, and the matrix equation $\boldsymbol{V'} = \boldsymbol{Z'I'}$ for the balanced system is given in eqn. (30).

The constraint matrix is built up from the following considerations. First, the currents in the f mesh are subject to the constraint of a line-to-ground fault on line a, so that

$$I_{f0} = I_{f1} = I_{f2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (58)$$

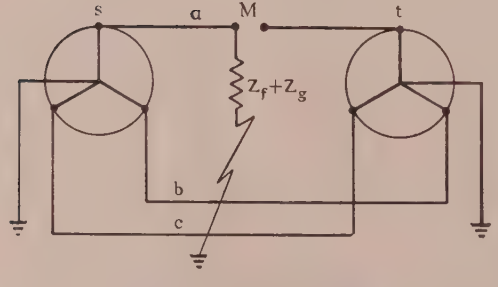


Fig. 18.—The system of Fig. 12, having an open conductor and simultaneous line-to-earth fault.

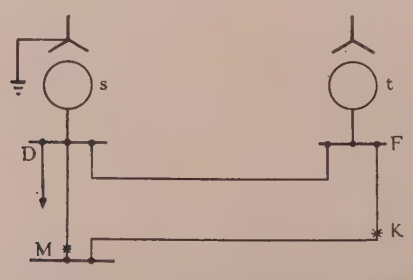


Fig. 19.—Power system to illustrate the general application of the technique.

D—Balanced star-load of Z_L ohms per phase.
M—Open conductor on line c, the end connected to D passing to earth through an arc impedance of Z_f and a total earth impedance of Z_g ohms.
K—Double line-to-earth short-circuit on lines a and b.

Secondly, the currents in the s mesh are subject to the constraint of an open conductor on line a, so that

$$I_{s0} + I_{s1} + I_{s2} = 0 \quad . \quad . \quad . \quad . \quad . \quad (59)$$

Thus we may analyse in terms of I_{f0}, I_{s0} and I_{s1}. The connection matrix $\boldsymbol{C}$ is then

$$\boldsymbol{C} = \begin{array}{c|ccc} \textit{Mesh} & f_0 & s_0 & s_1 \\ \hline f_0 & 1 & & \\ s_0 & & 1 & \\ f_1 & 1 & & \\ s_1 & & & 1 \\ f_2 & 1 & & \\ s_2 & & -1 & -1 \end{array} \quad . \quad . \quad . \quad (60)$$

The impedance matrix calculated from $\overline{C}_t \boldsymbol{Z'C}$ is found to be

$$\boldsymbol{Z''} = \begin{array}{c|ccc} \textit{Mesh} & f_0 & s_0 & s_1 \\ \hline f_0 & Z_{s0}+Z_{s1}+Z_{s2}+3(Z_g+Z_f) & Z_{s0}-Z_{s2} & Z_{s1}-Z_{s2} \\ s_0 & Z_{s0}-Z_{s2} & Z_{s0}+Z_{t0}+Z_{s2}+Z_{t2} & Z_{s2}+Z_{t2} \\ s_1 & Z_{s1}-Z_{s2} & Z_{s2}+Z_{t2} & Z_{s1}+Z_{t1}+Z_{s2}+Z_{t2} \end{array} \quad . \quad . \quad . \quad (61)$$

(3.3.5) **General Application.**

The technique demonstrated in the simple examples above has great power, for no further knowledge or ability is required to apply it to loaded systems with a number of faults occurring on any line. To show this, expressions for the voltage and impedance matrices will be derived for the more complicated system of Fig. 19.

The unbalanced faults are replaced by the appropriate balanced conditions as explained in the preceding Sections, and the equivalent sequence networks are deduced.

The zero- and positive-sequence networks are shown in Figs. 20 and 21, respectively. There is no need to draw the negative-sequence diagram, since it is the same as the positive one, except that the voltages are zero and the impedances have their negative-sequence values.

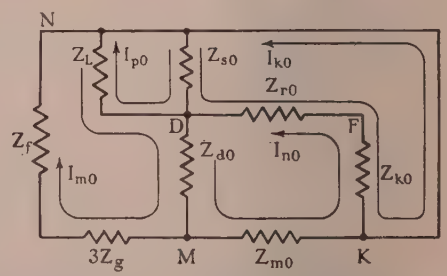


Fig. 20.—Zero-sequence network for the system of Fig. 19, after all unbalanced faults have been replaced by a balanced condition.

In choosing the paths of the mesh currents, which are the same in each sequence network, care should be taken to ensure that the fault constraint can be easily applied, by having each fault path traversed by one current only. Thus the current

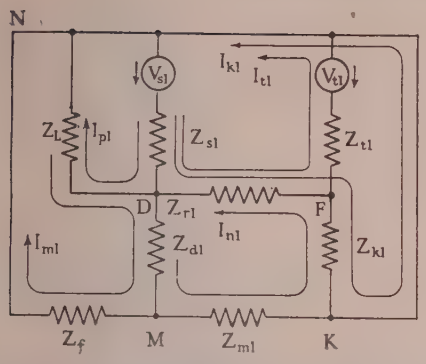


Fig. 21.—Positive-sequence network for the system of Fig. 19, after all unbalanced faults have been replaced by a balanced condition.

between D, M and K is subject to an open-conductor constraint, the current between D, M and N to a line-to-earth constraint, and the current from K to N to a double line-to-earth constraint. Furthermore, if the currents flowing through certain sections of the system are required, it is convenient if each of these sections is traversed by one mesh current only.

The mesh currents having been delineated the impedance matrices for each sequence diagram are then written down using Stigant's rule, and are

The impedance matrix of the balanced condition in terms of symmetrical components is then the compound diagonal matrix

$$\boldsymbol{Z}' = \begin{array}{|c|c|c|}\hline \boldsymbol{Z}'_{00} & & \\ \hline & \boldsymbol{Z}'_{11} & \\ \hline & & \boldsymbol{Z}'_{22} \\ \hline \end{array} \quad . \quad . \quad . \quad (65)$$

where $\boldsymbol{Z}'_{00}$, $\boldsymbol{Z}'_{11}$, and $\boldsymbol{Z}'_{22}$ have values given in eqns. (62), (63) and (64).

The corresponding voltage matrix is

$$\boldsymbol{V}' = (0, 0, 0, 0, V_{s1}, 0, V_{s1}, 0, V_{s1} - V_{t1}, 0, 0, 0, 0, 0) \quad . \quad (66)$$

The constraint matrix $\boldsymbol{C}$ is built up from the following considerations:

(*a*) The mesh currents I_p and I_t have no constraints imposed.

(*b*) I_n has the imposed constraint of an open conductor on phase *c*, i.e. $I_c = 0$, and $I_0 + aI_1 + a^2I_2 = 0$. The analysis may then proceed in terms of any two, say I_0 and I_2, and then

$$I_{n1} = -a^2I_{n0} - aI_{n2} \quad . \quad . \quad . \quad . \quad . \quad (67)$$

(*c*) The mesh current I_m has the imposed constraint of a line-to-earth fault on phase *c*, so that $I_a = I_b = 0$. Therefore

$$I_0 = \tfrac{1}{3}I_c, \quad I_1 = \tfrac{1}{3}a^2I_c, \quad I_2 = \tfrac{1}{3}aI_c$$

$$\boldsymbol{Z}'_{00} = \begin{array}{c|cccc} & k & n & p & m \\ \hline k & Z_{s0} + Z_{r0} + Z_{k0} & -Z_{k0} - Z_{r0} & Z_{s0} & \\ n & -Z_{k0} - Z_{r0} & Z_{d0} + Z_{m0} + Z_{k0} + Z_{r0} & & Z_{d0} \\ p & Z_{s0} & & Z_{s0} + Z_L & -Z_L \\ m & & Z_{d0} & -Z_L & Z_L - Z_{d0} + 3Z_g + Z_f \end{array} \quad . \quad . \quad . \quad . \quad . \quad (62)$$

$$\boldsymbol{Z}'_{11} = \begin{array}{c|ccccc} & k & n & p & m & t \\ \hline k & Z_{s1} + Z_{r1} - Z_{k1} & -Z_{k1} - Z_{r1} & Z_{s1} & & Z_{s1} + Z_{r1} \\ n & -Z_{k1} - Z_{r1} & Z_{d1} + Z_{m1} + Z_{k1} + Z_{r1} & & Z_{d1} & -Z_{r1} \\ p & Z_{s1} & & Z_{s1} + Z_L & -Z_L & Z_{s1} \\ m & & Z_{d1} & -Z_L & Z_{d1} + Z_L + Z_f & \\ t & Z_{s1} + Z_{r1} & -Z_{r1} & Z_{s1} & & Z_{s1} + Z_{r1} + Z_{t1} \end{array} \quad . \quad . \quad (63)$$

$$\boldsymbol{Z}'_{22} = \begin{array}{c|ccccc} & k & n & p & m & t \\ \hline k & Z_{s2} + Z_{r2} + Z_{k2} & -Z_{k2} - Z_{r2} & Z_{s2} & & Z_{s2} + Z_{r2} \\ n & -Z_{k2} - Z_{r2} & Z_{d2} + Z_{m2} + Z_{k2} + Z_{r2} & & Z_{d2} & -Z_{r2} \\ p & Z_{s2} & & Z_{s2} + Z_L & -Z_L & Z_{s2} \\ m & & Z_{d2} & -Z_L & Z_{d2} + Z_L + Z_f & \\ t & Z_{s2} + Z_{r2} & -Z_{r2} & Z_{s2} & & Z_{s2} + Z_{r2} + Z_{t2} \end{array} \quad . \quad . \quad . \quad (64)$$

Thus the analysis may proceed in terms of one current only, e.g. I_{m2}, so that

$$I_{m0} = a^2 I_{m2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (68)$$

and

$$I_{m1} = a I_{m2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (69)$$

(*d*) The mesh current I_k has the imposed constraint of a double line-to-earth fault on phase *a* and *b*, so that

$$I_0 + aI_1 + a^2 I_2 = 0$$

Analysing in terms of any two, say I_0 and I_2, we have

$$I_{k1} = -aI_{k2} - a^2 I_{k0} \quad . \quad . \quad . \quad . \quad . \quad (70)$$

The constraint matrix may now be built up and is

$$C = \begin{array}{|c|c|c|c|c|c|c|c|c|c|c|} \hline \textit{Sequence} & & 0 & & & & 1 & & & & 2 \\ \hline & \textit{Mesh} & k & n & p & p & t & k & n & p & m & t \\ \hline 0 & k & 1 & & & & & & & & & \\ \hline & n & & 1 & & & & & & & & \\ \hline & p & & & 1 & & & & & & & \\ \hline & m & & & & & & & & & a^2 & \\ \hline 1 & k & -a^2 & & & & & -a & & & & \\ \hline & n & & -a^2 & & & & & -a & & & \\ \hline & p & & & & 1 & & & & & & \\ \hline & m & & & & & & & & & a & \\ \hline & t & & & & & 1 & & & & & \\ \hline 2 & k & & & & & & 1 & & & & \\ \hline & n & & & & & & & 1 & & & \\ \hline & p & & & & & & & & 1 & & \\ \hline & m & & & & & & & & & 1 & \\ \hline & t & & & & & & & & & & 1 \\ \hline \end{array} \quad (71)$$

The ordering of the rows of this matrix is important; they must be ordered in the same way as the rows and columns of the impedance matrix $\boldsymbol{Z}'$. The ordering of the columns is not important when a computer solution is to be obtained, but where only a few currents are to be calculated by hand, using for example the reduction formulae of Kron (Reference 3, Chapter 10), it is necessary to place those desired currents in the columns on the left.

(3.4) Derivation of the Impedance Matrix from the Sequence Network

In those cases where the sequence network for the unbalanced fault condition can be drawn without undue difficulty, it is useful to be able to write down the voltage and impedance matrices immediately from the diagram. Stigant's rule enables this to be done.

The symmetrical-component representation of the problem of Section 3.3.1, with a line-to-earth fault on phase *b*, is shown in Fig. 22, which is obtained from Fig. 13, by using phase-

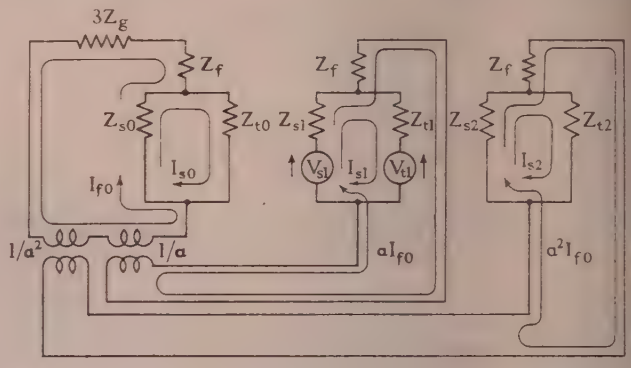


Fig. 22.—Symmetrical-component equivalent circuit of Fig. 12, for a line-to-earth fault at M on line *b* of impedance $Z_f + Z_g$ ohms.

shifting transformers to ensure the correct relationships between the sequence currents in the fault mesh. The meshes traversed in Fig. 22 are the same as those of Fig. 13, but the currents I_{f1} and I_{f2} are now expressed in terms of I_{f0}. There are then four independent currents I_{f0}, I_{s0}, I_{s1}, I_{s2}, and the impedance matrix is written down in terms of these currents and their respective mesh voltages.

Consider the mesh of the current I_{f0}. The voltage associated with the mesh is V_{s1}, but is separated from the current I_{f0} by the transformer of ratio $1 : a$. Thus the voltage of the f_0 mesh is

$$V_{s1}\frac{1}{a} = a^2 V_{s1}$$

The total impedance of the mesh is obtained by summing all the impedances through which I_f and its related currents flow, i.e. $3Z_g + 3Z_f + Z_{s0} + Z_{s1} + Z_{s2}$. There are never any a terms in the self-impedance of a mesh; for considering the impedance Z_{s2}, the voltage drop in this impedance for the current I_{f0} is $a^2 Z_{s2} I_{f0}$, and transferring this voltage through the transformer of ratio $1 : a^2$, the voltage drop becomes

$$a^2 Z_{s2} I_{f0}/a^2 = Z_{s2} I_{f0}$$

This cancellation of the phase shift must always take place when deriving the diagonal elements of the matrix.

The evaluation of the non-diagonal elements will be shown by finding two elements. The element (f_0, s_2) is found from the following considerations. The common impedance is Z_{s2} and both currents flow through it in the same direction. The voltage drop due to the current I_{s2} is $Z_{s2} I_{s2}$, which, on passing through the transformer of ratio $1 : a^2$, to the mesh of I_{f0} becomes $Z_{s2} I_{s2}/a^2 = aZ_{s2} I_{s2}$. Hence the element ($f_0$, s_2), is aZ_{s2}. The element (s_2, f_0) on the other hand represents the effect of the current I_{f0} on the mesh s_2. Once more the common impedance is $+Z_{s2}$, but since the actual current through the impedance is $a^2 I_{f0}$, the voltage drop in this impedance, due to the current I_{f0}, is $a^2 Z_{s2} I_{f0}$. Hence the element (s_2, f_0), is $a^2 Z_{2s}$.

The other elements of the matrix may be found in a similar manner giving the resultant impedance matrix

$$\mathbf{Z}'' = \begin{array}{c|c|c|c|c} & f_0 & s_0 & s_1 & s_2 \\ \hline f_0 & Z_{s1}+Z_{s2}+Z_{s0}+3Z_g+3Z_f & Z_{s0} & a^2Z_{s1} & aZ_{s2} \\ \hline s_0 & Z_{s0} & Z_{s0}+Z_{t0} & & \\ \hline s_1 & aZ_{s1} & & Z_{s1}+Z_{t1} & \\ \hline s_2 & a^2Z_{s2} & & & Z_{s2}+Z_{t2} \end{array} \quad . \quad . \quad (72)$$

which is the same as eqn. (39).

The impedance matrix of any interconnected sequence network can therefore be written down at sight regardless of inherent phase shifts.

(4) ACKNOWLEDGMENTS

The authors wish to express their appreciation of the encouragement they have received from Mr. J. A. Fitzpatrick of The General Electric Co., Ltd., and Mr. C. F. Partridge of the College of Technology, Birmingham. Acknowledgment is also made to the Directors of The General Electric Co., Ltd., for permission to publish the paper.

(5) REFERENCES

(1) MORTLOCK, J. R.: "The Evaluation of Simultaneous Faults on 3-phase Systems," *Journal I.E.E.*, 1947, **94**, Part II, p. 166.

(2) STIGANT, S. A.: "The Direct setting up of $Z_{\alpha\beta}$ for Closed Mesh Networks from the Network Diagram," *Beama Journal*, 1947, **54**, pp. 28 and 65.

(3) KRON, G.: "Tensor Analysis of Electrical Networks" (Wiley, 1939).

(4) MORTLOCK, J. R., and HUMPHREY DAVIES, M. W.: "Power System Analysis (Chapman and Hall, 1952).

(5) CLARK, E.: "Circuit Analysis of A.C. Power Systems 1. Symmetrical Components" (Wiley, New York, 1943; Chapman and Hall, London).

(6) WILLIAMS, F. C., ROBINSON, A. A., and KILBURN, T.: "Universal High-Speed Digital Computers—Serial Computing Circuits," *Proceedings I.E.E.*, Paper No. 1164 M, July, 1951 (**99**, Part II, p. 107).

(7) WILKES, M. V.: "Programme Design for High Speed Automatic Calculating Machines," *Journal of Scientific Instruments*, 1949, **26**, p. 217.

(8) WHEELER, D. J., and GILL, S.: "The Preparation of Programmes for an Electronic Digital Computer" (Addison-Wesley Press Inc., Massachusetts, U.S.A.).

621–526 : 621.373.421

Monograph No. 126 M
Apr. 1955

THE FREQUENCY-RESPONSE ANALYSIS OF NON-LINEAR SYSTEMS

By P. E. W. GRENSTED, M.A.

(*The paper was first received 2nd October, and in revised form 22nd December*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in April*, 1955.)

SUMMARY

If a non-linear system is oscillating, either through external excitation or internal regenerative action, it may be possible to show that the waveform at the input to the non-linear elements in the system is approximately sinusoidal. In such a case the frequency-response analysis may be carried out on the assumption that all harmonic components generated by the non-linear element can be ignored.

The paper discusses, with reference to feedback control systems, the results that can be obtained if this approximation is made. The amplitude, frequency and stability of steady self-excited oscillations are derived. Transient oscillations may be considered only in systems governed by second-order differential equations, but in these cases an analytic expression for the variation of frequency and damping with time, and hence for the full solution is derived. Several examples, all relating to "on-off" controllers, are given comparing the approximate and exact solutions; the accuracy obtained both for transient and steady oscillations is within about 10%.

In an Appendix the necessary extension to the familiar j notation is developed and used to derive a criterion of stability.

LIST OF SYMBOLS

A = Initial amplitude of error.
a = Amplitude of error.
a_s = Amplitude of error, steady-state solution.
$B = A - a_s$.
c = Coefficient of damping.
$\mathrm{D} = d/dt$.
e = Error.
$\mathrm{f}(e)$ = Non-linear characteristic.
$\mathrm{G}(\mathrm{D})$ = Transfer operator.
K = Magnitude of output of on-off element.
$\mathrm{n}(a) = \mathrm{p}(a) + j\mathrm{q}(a)$ = Describing function of gain.
$\mathrm{p}(a)$ = Describing function of in-phase gain.
$\mathrm{q}(a)$ = Describing function of quadrature gain.
T, T_1 and T_2 = Time lags.
T = Periodic time, accurate solution.
T_f = Periodic time, frequency-response solution.
$\alpha = \dot{a}/a$ = Negative damping factor.
$\lambda = \sqrt{(T_1/T_2)}$.
$\mu = -\alpha = -\dot{a}/a$ = Damping factor.
ϕ = Initial phase of error.
ψ = Phase of error.
$\omega = \dot{\psi}$ = Radial frequency.
ω_s = Radial frequency, steady-state solution.

(1) INTRODUCTION

Frequency-response methods may be used to analyse a non-linear system provided that the harmonics generated by the non-linear elements in the system may be ignored, for it may then be assumed that only one component of frequency is present throughout the system. A careful inspection of the attenuation of harmonics in the various loops of which the system is composed will indicate whether this approximation is justified. A block schematic is a useful way of describing the system for this purpose, and examples of a passive and dynamic system are given in Section 2. The general purpose of the paper is to show the results that may be attained if frequency-response methods are used.

Several authors[1–5] have applied frequency-response analysis to non-linear control systems, but other systems[6–9] have been handled with equal ease. In an early paper by Jacobsen[7] the forced oscillations of a passive mechanical system with one degree of freedom and non-linear damping are considered.

Previous workers have analysed the steady-state solution of non-linear systems. The amplitude, frequency and stability of self-excited oscillations, if any, are obtained for an isolated system and also the same results for the forced oscillations of an externally-excited system. The methods used are described in Section 3 with reference to the self-excited oscillations of a feedback controller. The accuracy that can be obtained is illustrated by two examples. Another example demonstrates how one apparently stable mode of oscillation can be swamped by another stable mode—a possibility which is usually overlooked.

In Section 4, the frequency-response analysis is extended so that transient solutions of non-linear systems may be obtained in an analytical form. The solution is obtained in terms of frequency and damping which are varying with time, and is restricted to systems governed by a second-order differential equation. Examples are given in Section 5 of the transient oscillations of an on-off controller, and the solutions obtained by the frequency-response analysis are compared with exact solutions.

An accuracy within about 10% is obtained in all the examples considered in the paper.

A formal analysis of a general non-linear system is attempted in Section 9 (Appendix), using the assumption that generated harmonics may be neglected. An operational notation, which is a generalized form of the familiar j notation, is used.

(2) BLOCK SCHEMATICS

Any dynamic or passive physical system may be represented by a block schematic. The output of each block in the schematic depends only on the signal applied to its input. In a linear system the mathematical relation which determines the output of a block in terms of its input is a linear differential equation in which all differentiations are with respect to time. This means that if the input magnitude is increased in a certain ratio, the output magnitude is increased in the same ratio. A non-linear system is defined as one in which this latter property does not hold for at least one of the blocks.

Blocks representing addition or subtraction are also required both for linear and non-linear systems.

By splitting the schematic down to blocks in their most elementary form it is usually possible to arrange that the non-linear blocks are independent of time and frequency. The output of such a block is an instantaneous function of its input. This is a desirable but not essential first step in the methods to be discussed.

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Grensted is in the Department of Engineering, University of Cambridge.

The importance of a block schematic is that it gives considerable insight into the physical behaviour of the system, and an indication of the relative importance of harmonics generated by any non-linearities which may be present.

Example (*a*).—A passive mechanical system consists of a mass M supported on a spring of stiffness K. The motion x of the mass is opposed by a frictional force P, where $P = P(v)$ is a non-linear function of the velocity v of the mass. If an external force $F(t)$ is applied to the mass, the block schematic for the system may be drawn as in Fig. 1.

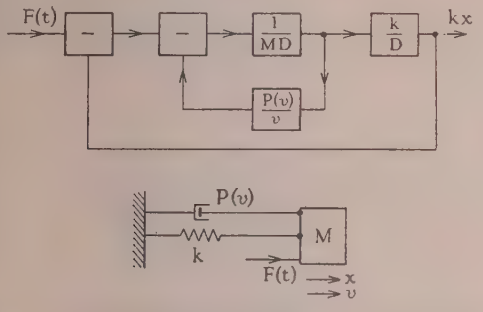


Fig. 1.—Block schematic of a non-linear mechanical system.

The notation used is that D is the operator d/dt, and the output of any block is given by multiplying the input by the operator or function written inside the block.

This system has been analysed by Jacobsen, using, in effect, the frequency-response methods of the paper, for the case where $F(t)$ is varying sinusoidally.

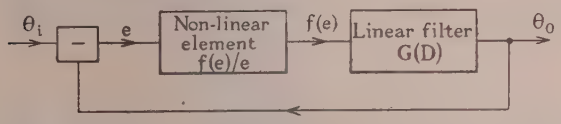


Fig. 2.—Block schematic of a non-linear feedback control system.

Example (*b*).—An example of a dynamic non-linear system is the feedback control shown in Fig. 2. This system is analysed in the remaining Sections.

(3) STABILITY AND STEADY-STATE SOLUTION OF A NON-LINEAR SYSTEM

This Section develops a method of investigating the stability of a non-linear system, and of determining the amplitude and frequency of steady oscillations.

(3.1) Application to Control System

The physical system considered is a feedback control system with a single frequency-independent non-linearity in the forward loop. The block schematic for the system is shown in Fig. 2.

The forward loop consists of a non-linear element whose output, $f(e)$, is an instantaneous function of the error, e, followed by the transfer operator $G(D)$ of a linear filter. If the input, θ_i, is zero the equation for the error is

$$G(D)f(e) + e = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

A steady-state solution of the form

$$e = a \sin \omega t \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

is looked for, where a is the steady-state amplitude and ω is the steady-state frequency. This expression for the error is substituted in eqn. (1) and $f(a \sin \omega t)$ is expanded into its Fourier components. Thus

$$G(D)(\Sigma a_n \sin n\omega t + b_0 + \Sigma b_n \cos n\omega t) + a \sin \omega t = 0 \quad . \quad (3)$$

A very considerable simplification in the analysis is made if b_0 is identically equal to zero. This occurs if $f(e)$ is such that its mean value is zero when e is varying in a sinusoidal manner. This restriction on $f(e)$ is now made, although it is not essential to the method.

(3.2) The Gain-Describing Function

The main approximation of the frequency-response analysis is made at this point. $G(D)$ must be the transfer operator of a low-pass filter, and its effect in eqn. (3) is to attenuate the harmonic terms generated by $f(e)$ relative to the fundamental term. The harmonic terms are thus neglected. Referring to the block schematic of Fig. 2 it is seen that it is assumed that any harmonics generated by the non-linear element are considered to be negligible by the time they return to the input of the element. In a system containing several non-linearities the generated harmonics must also be negligible before they reach any other non-linear element. Eqn. (3) now simplifies to

$$G(D)(a_1 \sin \omega t + b_1 \cos \omega t) + a \sin \omega t = 0 \quad . \quad (4)$$

where $$a_1 = \frac{1}{\pi}\int_0^{2\pi} \sin \psi f(a \sin \psi) d\psi$$

and $$b_1 = \frac{1}{\pi}\int_0^{2\pi} \cos \psi f(a \sin \psi) d\psi$$

a_1 and b_1 may be termed the in-phase and quadrature amplitude-describing functions of the non-linear element. They are expressions for the in-phase and quadrature components, respectively, of the amplitude of the fundamental component of the output signal from the element in terms of the amplitude of the input signal, a.

It is more convenient to consider the gain-describing function, $n(a)$, with components

$$p(a) = a_1/a = \frac{1}{\pi a}\int_0^{2\pi} \sin \psi f(a \sin \psi) d\psi \quad . \quad . \quad (5)$$

and $$q(a) = b_1/a = \frac{1}{\pi a}\int_0^{2\pi} \cos \psi f(a \sin \psi) d\psi \quad . \quad . \quad (6)$$

Eqn. (4) may now be written

$$G(j\omega)n(a) + 1 = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where $$n(a) = p(a) + jq(a) \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Eqn. (7) is that which would be obtained if the system were a linear one with a constant loop gain n. When the gain is non-linear, n becomes a function of the amplitude of the signal at the input of the non-linearity, as shown by eqns. (5) and (6). There may also be a phase shift across the non-linearity dependent on amplitude alone. Values of a and ω satisying eqn. (7) represent the amplitude and frequency of possible modes of oscillation of the system.

One method of finding values of a and ω satisfying eqn. (7) is by an extension of the normal Nyquist diagram. The locus of $G(j\omega)$ with ω as a parameter is plotted in the complex plane. Also $-1/n(a)$, with a as a parameter, is plotted in the same plane. Curves shown in Fig. 3 are a typical example. Intersections of the two curves as at A, B and C give solutions

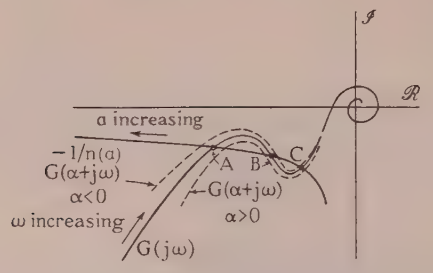


Fig. 3.—Nyquist diagram for a non-linear control system.

(ω_A, a_A), (ω_B, a_B) and (ω_C, a_C) for steady-state frequencies and amplitudes.

(3.3) Stability of Oscillations

It is now necessary to investigate the stability of the possible modes of oscillation. As well as plotting $G(j\omega)$ it is useful to plot or visualize curves of $G(\alpha + j\omega)$, with α constant. Now a solution of the more general form of eqn. (7),

$$G(\alpha + j\omega)n(a) + 1 = 0 \quad . \quad . \quad . \quad . \quad (9)$$

corresponds to an oscillation of frequency ω and amplitude a, which is varying exponentially with time with a damping factor $-\alpha$. For a linear system this statement is exactly true; for a non-linear system it is only approximately true, even if $|\alpha| \ll \omega$. The degree of approximation involved is discussed in Section 9 (Appendix).

Now since $G(\alpha + j\omega)$ is a conformal transformation from the $(\alpha + j\omega)$ plane, lines of $G(\alpha + j\omega)$ with α or ω constant may be mapped directly from the plot of $G(j\omega)$ in the Nyquist diagram. In particular, if α is a constant greater than zero, lines of $G(\alpha + j\omega)$ will be on the right-hand side of the $G(j\omega)$ locus when this locus is followed in the direction of ω increasing; if α is a constant less than zero, lines of $G(\alpha + j\omega)$ will lie on the left-hand side of the $G(j\omega)$ locus.

Consider the stability of the point A in Fig. 3. If a small disturbance causes an increase in amplitude, α becomes negative, and hence the oscillations become positively damped. A small decrease in amplitude results in negative damping. Thus the oscillations at A are stable. Similar reasoning shows that the oscillations at B and C are unstable and stable respectively.

In general, it may be stated that an intersection of the two curves represent a stable mode of oscillation if the direction of the $-1/n(a)$ locus is turned anti-clockwise from the direction of the $G(j\omega)$ locus, and an unstable mode in the converse case.

In Section 9 (Appendix) it is shown that eqn. (9) may be inadequate to describe the behaviour of the oscillation, even in the neighbourhood of a steady-state solution, and that a slightly different criterion of stability should be used. Eqn. (9) is based on the unjustifiable assumption that the derivatives of α and ω are small compared with α^2 and $(\omega - \omega_s)^2$. However, the criterion given in the text is simpler to apply, and appears to give the correct solution for all normal types of feedback control systems.

(3.4) Example

The preceding analysis of the stability of steady oscillations must be used with caution when applied to a system which has more than one possible mode of oscillation. The following example illustrates this point.

Example (c).—Consider a system with a non-linear element of the on-off type and a transfer operator representing a finite time lag, T, followed by an integration.

Then $$G(D) = \varepsilon^{-TD}/D$$

and $$f(e) = K, \text{ if } e > 0$$

$$= -K, \text{ if } e < 0$$

The gain-describing function has no quadrature component and is simply

$$n(a) = \frac{4}{\pi}\frac{K}{a}$$

Hence the locus of $-1/n(a)$ lies along the negative real axis in the Nyquist diagram. The locus of the transfer function $G(j\omega) = (\cos \omega T - j \sin \omega T)/j\omega$ spirals towards the origin in a clockwise manner as ω increases, as shown in Fig. 4.

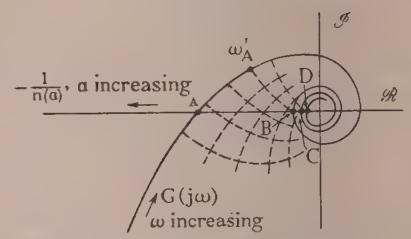


Fig. 4.—Nyquist diagram [example (c)].

Conditions for steady oscillations thus occur when $G(j\omega)$ intersects the negative real axis, i.e. when

$$\mathscr{I}[G(j\omega)] = 0, \text{ and } \mathscr{R}[G(j\omega)] < 0$$

Steady-state frequencies obtained by the frequency-response analysis are thus given by

$$\omega_s = \pi/(2T), 5\pi/(2T), 9\pi/(2T), \text{ etc.}$$

From the reasoning given in Section 3.3 all these values of ω would appear to represent stable modes. However, let us consider steady oscillations taking place at the point B in Fig. 4 with frequency $\omega_B = 5\pi/2T$. Although the damping, $-\alpha$, corresponding to this point is zero, and is such as to maintain stable oscillations at this point, the value of α corresponding to some frequency ω'_A near the lowest natural frequency $\omega_A = \pi/2T$ is large and positive. Thus a component of frequency ω_A will eventually build up, and the oscillation at B breaks down until steady conditions at A are reached. At A, however, the damping associated with the point B is negative. It follows that the lowest natural frequency, ω_A, is the only stable mode of oscillation possible, since oscillations at C, D, etc., will break down in a similar manner. This possibility of one mode of oscillation swamping another is characteristic of non-linear systems, and is usually overlooked in descriptions of the frequency-response method.

An exact solution for this system is easily obtained for comparison. From eqn. (1) the equation for the error is

$$\dot{e} = \delta K$$

where δ becomes $+1$ at a time T after e becomes negative, and -1 at a time T after e becomes positive.

Solutions for steady-state oscillations are shown in Fig. 5. Fig. 5(*a*) shows the lowest frequency possible. Changes in velocity $\dot{e}$ occur at a', b', c', etc., as a result of changes in the sign of e at a, b, c, etc. The periodic time is $4T$. The next higher frequency is shown in Fig. 5(*b*), and has a periodic time of $4T/5$. In Fig. 5(*c*) it is demonstrated that this mode of oscillation is

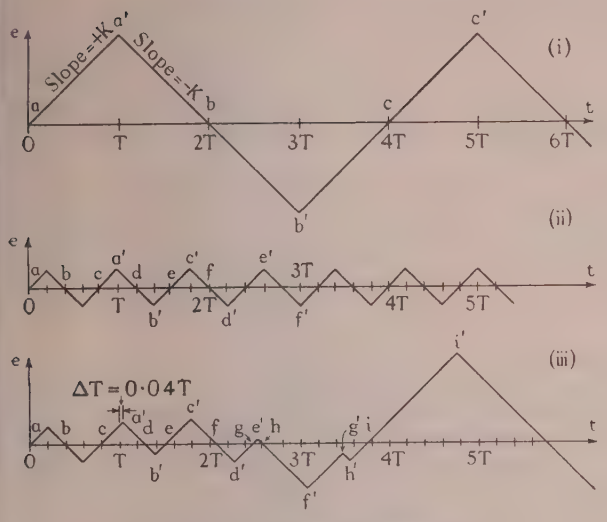


Fig. 5.—Steady oscillations [example (*c*)].
(i) First mode. Stable.
(ii) Second mode. Unstable.
(iii) Breakdown of second mode.

unstable and breaks down when a slight disturbance occurs. If the change in slope at a' occurs a small time ΔT late, the resulting waveform is as shown, and very soon becomes that of the lowest frequency.

It is thus verified that only the lowest natural frequency is stable. It may be noticed that the frequencies obtained by the frequency-response analysis in this example are identical with those obtained by an exact analysis.

(3.5) Accuracy of Method

It is difficult to obtain any simple criterion for the accuracy of the method. The approximation made is that harmonics generated by the non-linear element are neglected, and hence the accuracy depends on the amount of harmonic distortion generated by the non-linearity and the relative attenuation of the fundamental and harmonics round the loop.

Johnson[3] gives an iterative method for taking the third harmonic into consideration; the change in the calculated steady-state frequency thereby introduced gives an indication of the accuracy achieved.

A general guide to the accuracy of the method may be obtained by considering some particular examples for which an exact solution may be derived. The following two examples relate to a control in which the non-linearity is an on-off element, which has an output $\pm K$ depending only on the sign of the error.

Example (*d*).—G(*D*) consists of two time delays, T_1 and T_2, and one integration.

Thus $$G(D) = \frac{1}{1 + T_1 D} \frac{1}{1 + T_2 D} \frac{1}{D}$$

As in the example of Section 3.4, stable steady oscillations occur when $\mathscr{I}[G(j\omega)] = 0$, i.e. when $T_1 T_2 \omega^2 = 1$. The periodic time T_f, found by the frequency-response analysis, is thus

$$T_f = 2\pi\sqrt{(T_1 T_2)}$$

From eqn. (1) the true equation for the system may be written in the form

$$T_1 T_2 \dddot{e} + (T_1 + T_2)\ddot{e} + \dot{e} + K\,(\text{sign } e) = 0$$

and can be solved in a piece-wise manner. By equating conditions at the beginning and end of a cycle it is found that the true periodic time T must satisfy the equation

$$T = \frac{4}{T_2 - T_1}\left[T_2^2 \tanh\left(\frac{T}{4T_2}\right) - T_1^2 \tanh\left(\frac{T}{4T_1}\right)\right]$$

The solution is $T = k\sqrt{(T_1 T_2)}$, where k is a function of $\lambda = \sqrt{(T_1/T_2)}$ and varies between $4\sqrt{3} = 2\pi \times 1{\cdot}101$ when $\lambda = 0$ or ∞, and $2\pi \times 1{\cdot}022$ when $\lambda = 1$.

Curves of $T_f/\sqrt{(T_1 T_2)}$ and $T/\sqrt{(T_1 T_2)}$ against λ are shown in Fig. 6.

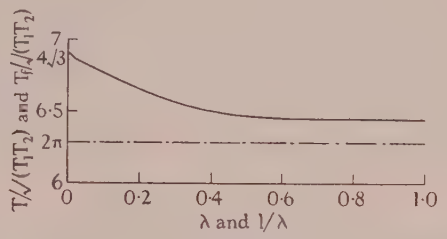


Fig. 6.—Comparison of periodic times of steady oscillations [example (*d*)].
– – – – – Frequency-response method: $T_f/\sqrt{(T_1 T_2)}$.
——— Exact analysis: $T/\sqrt{(T_1 T_2)}$.
$\lambda = \sqrt{(T_1/T_2)}$.

Example (*e*).—G(D) consists of a finite time lag T_1, a time delay T_2 and an integration.

Thus $$G(D) = \frac{1}{\varepsilon^{T_1 D}} \frac{1}{1 + T_2 D} \frac{1}{D}$$

Steady oscillations occur when

$$\mathscr{I}[G(j\omega)] = 0$$

i.e. when $\omega T_2 \tan \omega T_1 = 1$.

The lowest root of this equation gives the only stable mode of oscillation. The periodic time, T_f, found by the frequency-response analysis may thus be written $T_f = k_f\sqrt{(T_1 T_2)}$, where k_f is a function of $\lambda = \sqrt{(T_1/T_2)}$, and $k_f \to 2\pi$ as $\lambda \to 0$ and $k_f \to 4(\lambda + 1/\lambda)$ as $\lambda \to \infty$.

The equation for the true period time T is found, by the method outlined in example (*d*), to be

$$8T_2 \varepsilon^{T_1/T_2} = (4T_2 + 4T_1 - T)[1 + \varepsilon^{T/(2T_2)}]$$

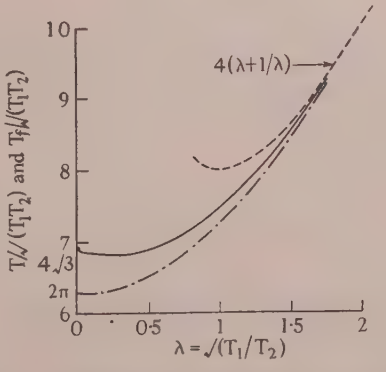


Fig. 7.—Comparison of times of steady oscillations. [example (*e*)].
– – – – – Frequency-response method: $T_f/\sqrt{(T_1 T_2)}$.
——— Exact analysis: $T/\sqrt{(T_1 T_2)}$.
$\lambda = \sqrt{(T_1/T_2)}$.

and the solution is written as $T = k\sqrt{(T_1T_2)}$, where k is a function of λ:

$$k \to 4\sqrt{3} \text{ as } \lambda \to 0, \text{ and } k \to 4(\lambda + 1/\lambda) \text{ as } \lambda \to \infty$$

Curves of $T_f/\sqrt{(T_1T_2)}$ and $T/\sqrt{(T_1T_2)}$ against λ are shown in Fig. 7.

In both these examples it is seen that the greatest error in the steady-state frequency found by the frequency-response analysis is about 10%.

(4) TRANSIENT OSCILLATIONS OF A NON-LINEAR SYSTEM

The methods of the previous Section are now used to obtain the transient response of non-linear systems which give rise to second-order differential equations. An analytical solution is obtained which is a good approximation to the true solution for all values of time.

Meanings are given to the terms "frequency" and "damping" when these are varying with time.

(4.1) Derivation of Equation

The system of Fig. 2 is investigated with G(D) the transfer operator of a velocity-lag servo mechanism. It is convenient to consider time and error as dimensionless quantities. This may be done without loss of generality, and we can write

$$G(D) = \frac{1}{2c + D}\,\frac{1}{D} \quad . \quad . \quad . \quad . \quad . \quad (10)$$

and $f(e)$ is equal to unity at some pertinent point of its characteristic.

If a solution $e(t)$ is obtained to this system, with $c = 1/2$, another system having a transfer operator

$$G'(D) = \frac{1}{1 + T'D'}\,\frac{1}{T'D'} \quad . \quad . \quad . \quad . \quad . \quad (11)$$

and a non-linear characteristic

$$f(e') = kf(e'/k) \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

has a solution

$$e'(t') = ke(t/T) \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

Primed symbols denote quantities and functions in the new system.

From eqns. (1) and (10)

$$\ddot{e} + 2c\dot{e} + f(e) = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

In order to obtain the solution of the more general system defined by eqns. (11) and (12) we need only consider the case $c = 1/2$ in eqn. (14). However, c is retained as a parameter for the present so that a comparison may be made with the familiar linear equation

$$\ddot{e} + 2c\dot{e} + b^2e = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

Eqn. (14) is the non-linear differential equation which we shall attempt to solve.

(4.2) Equations for Frequency and Damping

An oscillatory solution of the form

$$e = a \sin \psi \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

is assumed, in which the amplitude a and the phase ψ are functions of time t.

The frequency ω of this solution is defined as the rate of change of phase.

Thus $$\omega = \dot{\psi}; \text{ or } \psi = \int \omega dt \quad . \quad . \quad . \quad . \quad (17)$$

The damping, μ, of the solution is defined as the rate of reduction in amplitude divided by the amplitude.

Thus $$\mu = -\frac{\dot{a}}{a}; \text{ or } a = \varepsilon^{-\int \mu dt} \quad . \quad . \quad . \quad . \quad (18)$$

e may now be written in terms of frequency and damping as

$$e = A\varepsilon^{-\int_0^t \mu dt} \sin\left(\int_0^t \omega dt + \phi\right) \quad . \quad . \quad . \quad (19)$$

where A and ϕ are arbitrary constants.

If eqn. (19) is substituted in the original eqn. (14), the expansion of $f(e)$ will provide harmonic terms which are neglected for exactly the same reason as given in Section 3.2. The fact that ω is no longer constant does not alter the argument justifying this step, and the same order of accuracy may be expected for transient solutions obtained by using this approximation as was found for steady-state solutions, namely 10%.

The substitution may be done directly in eqn. (14), or by using the operational method developed in Section 9 (Appendix) for a general transfer operator of the form $G(D) \equiv R(D)/S(D)$, where $R(D)$ and $S(D)$ are polynomials in D. If the latter method is used, we have immediately

$$n(a) + (2c + D - \mu + j\omega)(D - \mu + j\omega) = 0$$

Hence $$\omega^2 = p(a) - 2c\mu + \mu^2 - \dot{\mu} \quad . \quad . \quad . \quad . \quad (20)$$

$$-\frac{\dot{a}}{a} = \mu = c + \frac{1}{2}\frac{\dot{\omega}}{\omega} + \frac{1}{2}\frac{q(a)}{\omega} \quad . \quad . \quad . \quad . \quad . \quad (21)$$

where the gain, $n(a) = p(a) + jq(a)$, is defined by eqns. (5) and (6).

Eqns. (20) and (21) are the basic equations from which μ and ω may be found when the frequency-response method is applied to a velocity-lag servo mechanism, or a non-linear equation of the form of eqn. (14). The equations are themselves non-linear, but a good approximation to their solution may be obtained by a simple iterative process.

(4.3) Comparison with Solution of Linear Equation

If the system is linear, i.e. $f(e) = b^2e$, then $p(a) = b^2$ and $q(a) = 0$. Eqns. (20) and (21) are then satisfied if $\dot{\mu} = \dot{\omega} = 0$, and $$\mu = c; \omega^2 = b^2 - c^2$$

These are the familiar expressions for damping and frequency of the linear eqn. (15).

(4.4) Comparison with Other Methods

If eqns. (20) and (21) had been derived from the approximate solution given by eqn. (9), identical equations would result, but with the terms involving the rates of change of damping and frequency missing. Examples given later show that these terms can be significant, and in general the more accurate eqns. (20) and (21) must be used.

Krylov and Bogoluibov[10] use a similar method of obtaining transient solutions to equations of the form of eqn. (14). In the notation of this paper they derive equations for μ and ω as

$$\left.\begin{aligned} 2\omega\nu - \nu^2 &= p(a) \\ \mu &= c + \frac{1}{2}\frac{q(a)}{\nu} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (22)$$

where ν, which is a constant, is a close approximation to the frequency and is estimated by inspection of the original equation. These authors obtain better approximations by considering higher harmonics.

Eqns. (22) are only of value when the damping is small and $f(e)$ is nearly linear.

(4.5) Solution of Fundamental Equations

The methods of solution of the fundamental eqns. (20) and (21) are now discussed.

Steady-state oscillations occur when $\mu = \dot{\omega} = 0$. Thus from eqns. (20) and (21) it is seen that

$$\left.\begin{aligned} \omega_s^2 &= p(a_s) \\ 0 &= c + \frac{1}{2}\frac{q(a_s)}{\omega_s} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (23)$$

where a_s and ω_s are the amplitude and frequency, respectively, of steady-state oscillations.

To obtain the transient values of a and ω various approximate or iterative methods must be used. In general these depend on the form of transient expected. It is convenient to discuss the cases where $q(a) \equiv 0$ and $q(a) \not\equiv 0$ separately.

(4.5.1) $q(a) \equiv 0$.

If $q(a) \equiv 0$, from eqn. (21) it follows that no steady-state oscillations of finite frequency or amplitude are possible. $q(a) \equiv 0$ whenever $f(x)$ is a single-valued odd function of x.

Eqn. (21) becomes

$$-\frac{\dot{a}}{a} = \mu = c - \frac{1}{2}\frac{\dot{\omega}}{\omega} \quad . \quad . \quad . \quad . \quad . \quad (24)$$

which can be integrated to give

$$a = A\varepsilon^{-ct}\sqrt{\frac{\omega(0)}{\omega}} \quad . \quad . \quad . \quad . \quad . \quad (25)$$

where A and $\omega(0)$ are the initial values of amplitude and frequency, respectively.

Now the term $\dot{\omega}/(2\omega)$ in eqn. (24), or the factor $\sqrt{(\omega(0)/\omega)}$ in eqn. (25), can be considered as a correction made necessary by the non-linearity, and approximate values of ω and $\dot{\omega}$ may be used for these terms. If c^2, μ^2, $\dot{\mu} \ll p(a)$, then an approximate value for ω is, from eqn. (20),

$$\omega_1 = \sqrt{[p(a)]} \quad . \quad . \quad . \quad . \quad . \quad (26)$$

Substituting ω_1 for ω in eqn. (25) gives the following equation for amplitude in terms of time t:

$$a = A\varepsilon^{-ct}\left[\frac{p(A)}{p(a)}\right]^{1/4} \quad . \quad . \quad . \quad . \quad (27)$$

Having thus obtained a relation between amplitude and time, a more accurate value of ω may be obtained from eqn. (20) using the fact that $\mu = -\dot{a}/a$.

The solution thus obtained should be sufficiently accurate for most purposes, but the iteration process can be repeated if better accuracy is required. One difficulty lies in the fact that if $p(a)$ is a complicated function of a, the amplitude as a function of time can only be obtained graphically from eqn. (27).

Once ω and a have been found as functions of time, the complete solution

$$e = a\sin\left(\int_0^t \omega dt + \phi\right)$$

may be formed. The constants A and ϕ are chosen so that the initial conditions are satisfied.

(4.5.2) $q(a) \not\equiv 0$.

$q(a)$ is finite when $f(e)$ is not a single-valued function of e. Such cases occur in systems which contain hysteresis, backlash or other hereditary phenomena. The value of $f(e)$ then depends on the sign of $\dot{e}$ as well as e. $q(a)$ may also be finite when $f(e)$ is considered as a function of frequency ω, as well as amplitude. An example is where $f(e)$ contains a finite time lag as well as a single-valued non-linearity. This case will be illustrated in an example later.

If $q(a)$ is finite it is not possible to integrate eqn. (21) directly. However, if we write

$$-\frac{\dot{a}}{a} = \mu = c + \frac{1}{2}\frac{\dot{\omega}_1}{\omega_1} + \frac{1}{2}\frac{q(a)}{\omega_1}$$

where $$\omega_1 = \sqrt{[p(a)]}$$

a first-order differential equation is derived:

$$-\dot{a} = a\frac{c + \frac{1}{2}\sqrt{\frac{q(a)}{p(a)}}}{1 + \frac{1}{4}\frac{ap'(a)}{p(a)}} \quad . \quad . \quad . \quad . \quad (28)$$

Although eqn. (28) is an approximate solution, it does give the correct value for a_s.

The solution of eqn. (28) is

$$t = \int_a^A \left[a\frac{c + \frac{1}{2}\sqrt{\frac{q(a)}{p(a)}}}{1 + \frac{1}{4}\frac{ap'(a)}{p(a)}}\right]^{-1} da$$

This relation between amplitude and time enables a more accurate value of ω to be obtained from eqn. (20).

However, the technique of solving eqns. (20) and (21) depends very much on the form of $f(e)$ and the initial amplitude. A few examples will now be worked out and their solutions compared with the exact solutions to eqn. (14) obtained by other means.

(5) EXAMPLES OF TRANSIENT SOLUTION

All the examples to be discussed are connected with simple on-off control systems. The non-linearities considered are characteristic of, for instance, a polarized relay. The transfer operator is of the form discussed in the previous Section.

(5.1) Idealized Relay

Example (f).—The case of an idealized relay is considered first, with no hysteresis, dead zone or time lag. Then $f(e)$ is given by the simple relation

$$f(e) = 1, \; e > 0$$
$$f(e) = -1, \; e < 0$$

and the equation to be solved is

$$\ddot{e} + 2c\dot{e} + e/|e| = 0$$

The initial conditions considered are

$$e = 1, \; \dot{e} = 0 \text{ at } t = 0$$

From eqn. (5), $p(a) = (4/\pi)(1/a)$, and since $f(e)$ is a single-valued function of e, $q(a) = 0$.

Eqn. (27) gives a first approximation for a as a function of time:

$$a = A\varepsilon^{-ct}\left(\frac{a}{A}\right)^{1/4}$$

i.e. $a = A\varepsilon^{-4ct/3}$

whence $\mu = \frac{4}{3}c \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$

In this example it appears that to a first approximation the damping is constant, and the presence of the non-linearity increases its value by one-third from the value it would have if the system were linear.

The following expression for ω is found by substituting this value of μ into eqn. (20):

$$\omega = \sqrt{\left(\frac{4}{\pi}\frac{1}{A}\varepsilon^{4ct/3} - \frac{8}{9}c^2\right)}$$

or $$\omega \simeq \sqrt{\left(\frac{4}{\pi}\frac{1}{A}\right)}\varepsilon^{2ct/3} - \frac{1}{2}\frac{8}{9}c^2\sqrt{\left(\frac{\pi A}{4}\right)}\varepsilon^{-2ct/3}, \text{ if } Ac^2 \ll 1 \quad . \quad . \quad . \quad . \quad (30)$$

Substituting eqns. (29) and (30) in eqn. (19) gives the solution

$$e = A\varepsilon^{-4ct/3}\sin\left[\frac{3}{2(c)}\sqrt{\left(\frac{4}{\pi}\frac{1}{A}\right)}(\varepsilon^{2ct/3} - 1) + \frac{2}{3}c\sqrt{\left(\frac{\pi}{4}A\right)}(\varepsilon^{-2ct/3} - 1) + \phi\right]$$

The initial conditions require that

$$\left.\begin{aligned} A\sin\phi &= 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \\ \tan\phi &= \frac{\omega(0)}{\mu(0)} = \left[\sqrt{\left(\frac{4}{\pi}\frac{1}{A}\right)} - \frac{1}{2}\frac{8}{9}c^2\sqrt{\left(\frac{\pi}{4}A\right)}\right]\Big/\left(\frac{4}{3}c\right) \end{aligned}\right\} (31)$$

For $c = \frac{1}{2}$, eqns. (31) gives $A = 1{\cdot}25$, $\phi = 53°$; and so the complete solution becomes

$$e = 1{\cdot}25\varepsilon^{-2t/3}\sin\left[173\varepsilon^{t/3} + 19\varepsilon^{-t/3} - 139\right]° \quad . \quad (32)$$

Eqn. (32) is shown in Fig. 8. Positions of the zeros and maxima of the true solution of the original equation are also

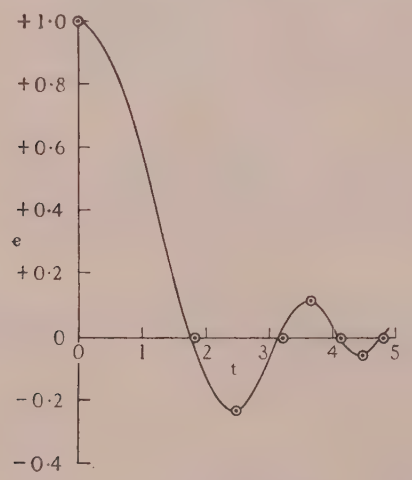


Fig. 8.—Transient solution. On-off controller with idealized relay [example (*f*)].

○ Peaks and zeros of accurate solution.

shown. Considering the fundamental approximation made in neglecting harmonics throughout, the agreement is as good as or even better than could be expected.

(5.2) Backlash

Example (*g*).—The case of a relay with a backlash or hysteresis effect is now considered. The form of f(*e*) is shown in Fig. 9.

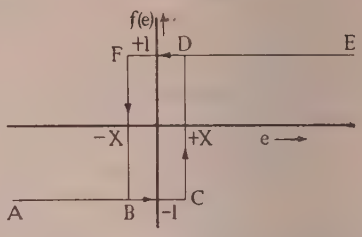


Fig. 9.—Characteristic of on-off element with backlash.

If e is increasing, f(e) follows the path ABCDE, and if e is decreasing f(e) follows the path EDFBA. The width of the backlash zone is $2X$.

The same initial conditions as in the previous example are chosen: $e = 1$, $\dot{e} = 0$ at $t = 0$.

From eqns. (5) and (6) the in-phase and quadrature gains become, respectively,

$$\mathrm{p}(a) = \frac{4}{\pi}\frac{1}{a}\sqrt{[1 - (X/a)^2]}$$

$$\mathrm{q}(a) = -\frac{4}{\pi}\frac{1}{a}(X/a)$$

The general form of the transient may be found from the Nyquist diagram. Oscillations tend to a steady condition. The steady-state amplitude is determined from eqn. (23), and thus

$$c - \frac{1}{2}\frac{4}{\pi}\frac{1}{a_s}\frac{X/a_s}{\sqrt{[1 + (X/a_s)^2]}} = 0 \quad . \quad . \quad . \quad (33)$$

In order to simplify the arithmetic and to give a representative example, a backlash zone is chosen such that the steady-state amplitude is a round figure. When $c = 1/2$, eqn. (33) gives $a_s = 0{\cdot}10$ when $X = 0{\cdot}0274$.

Substituting $X = 0{\cdot}0274$ and $c = 1/2$ in eqn. (28) a relation between $\dot{a}$ and a is obtained. This relationship is very close to a linear one, and leads to the conclusion that a may be written in the form

$$a = B\varepsilon^{-\beta t} + a_s \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

where B is an arbitrary constant, and β is the mean value of $-\dot{a}/(a - a_s)$ over the range considered; i.e. $\beta = 0{\cdot}70$ in this example.

Eqn. (34) becomes the first approximation for amplitude.

To obtain an approximation for frequency, eqn. (20) is used, with the value of μ derived from eqn. (34).

Thus $$\mu = -\frac{\dot{a}}{a} = \left(1 - \frac{a_s}{a}\right)\beta$$

$$\dot{\mu} = -\frac{a_s}{a}\left(1 - \frac{a_s}{a}\right)\beta^2$$

Hence $$\omega = \sqrt{\left[\mathrm{p}(a) - (2c - \beta)\left(1 - \frac{a_s}{a}\right)\beta\right]} \quad . \quad . \quad (35)$$

The method of iteration adopted ensures that the solution gives the correct values of steady-state frequency and amplitude.

The full solution becomes

$$e = a\sin\left(\int_0^t \omega dt + \phi\right) \quad . \quad . \quad . \quad . \quad (36)$$

where a and ω are given by eqns. (34) and (35) respectively, and the constants B and ϕ are determined by the initial conditions.

$$B = 1{\cdot}12 \text{ and } \phi = 55°$$

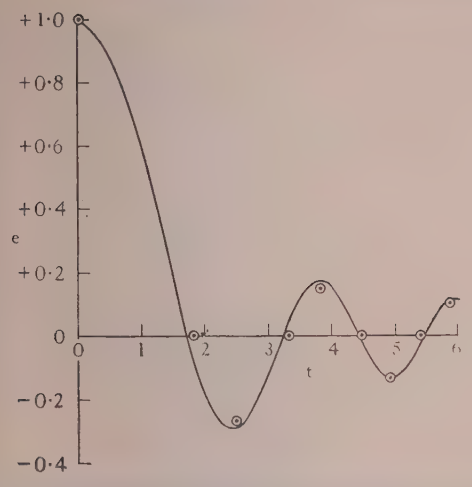


Fig. 10.—Transient solution. On-off controller with backlash in relay [example (*g*)].

⊙ Peaks and zeros of accurate solution.

The full solution [eqn. (36)] is shown in Fig. 10 and compared with the true solution of the original equation.

(5.3) Finite Time Delay

Example (*h*).—This example concerns an idealized relay with a finite time delay in its operation. The non-linearity is thus a function of frequency as well as amplitude.

The form of f(*e*) is now f(*e*) = δ, where δ = 1 at a time T after e becomes positive, and δ = − 1 at a time T after e becomes negative. This may be regarded as a phase lag λ proportional to frequency, with $\lambda = \omega T$.

Hence, the expressions for p(*a*) and q(*a*) are

$$\mathrm{p}(a) = \frac{4}{\pi}\frac{\cos \omega T}{a}$$

$$\mathrm{q}(a) = -\frac{4}{\pi}\frac{\sin \omega T}{a}$$

Eqns. (20) and (21) now become

$$-\frac{\dot{a}}{a} = \mu = c + \frac{1}{2}\frac{\dot{\omega}}{\omega} - \frac{1}{2}\frac{4}{\pi}\frac{\sin \omega T}{a\omega} \quad . \quad . \quad (37)$$

and

$$\omega^2 = \frac{4}{\pi}\frac{\cos \omega T}{a} - 2c\mu + \mu^2 - \dot{\mu} \quad . \quad . \quad . \quad (38)$$

The steady-state frequency ω_s is found from eqns. (37) and (38) when $\mu = \dot{\mu} = \dot{\omega} = 0$. If $c = 1/2$,

$$\omega_s \tan \omega_s T = 1 \quad . \quad . \quad . \quad . \quad . \quad (39)$$

A value 0·08 is chosen for the time lag T. Then from eqn. (39)

$$\omega_s = 3{\cdot}54$$

and from eqn. (38) $\quad a_s = 0{\cdot}10$

The initial conditions are again taken as $e = 1$, $\dot{e} = 0$ at $t = 0$.

The transient will then be one of increasing frequency, until the frequency ω_s is reached. Thus $\omega T < \omega_s T = 0{\cdot}283$. In order to solve eqns. (37) and (38) it therefore is assumed that

$$\cos \omega T = 1 \text{ and } \sin \omega T = \omega T$$

As far as the correction terms in eqn. (37) are concerned we may write

$$\omega^2 = \frac{4}{\pi}\frac{1}{a}, \text{ or } \frac{\dot{\omega}}{\omega} = -\frac{1}{2}\frac{\dot{a}}{a}$$

and eqn. (37) becomes

$$3\dot{a} = -4ca + 2a_s$$

with solution $\quad a = a_s + Be^{-4ct/3} \quad . \quad . \quad . \quad . \quad . \quad (40)$

As in the previous example the value of μ from eqn. (40) is substituted in eqn. (38). Cos ωT is still considered as unity. Then

$$\omega = \sqrt{\left[\frac{4}{\pi}\frac{1}{a} - \frac{8}{9}c^2\left(1 - \frac{a_s}{a}\right)\right]} \quad . \quad . \quad . \quad (41)$$

With the given initial conditions and $c = 1/2$, the initial amplitude A is found to be 1·21, and the initial phase $\phi = 55{\cdot}7°$.

The full solution based on eqns. (40) and (41) is shown in Fig. 11 and compared with the true solution of the original equation.

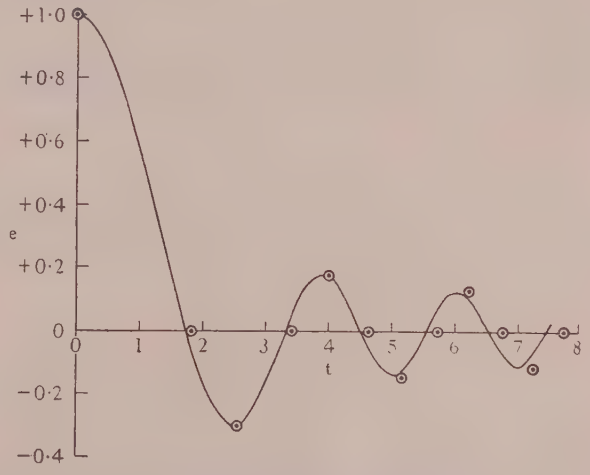


Fig. 11.—Transient solution. On-off controller with time delay in relay [example (*h*)].

⊙ Peaks and zeros of accurate solution.

(6) CONCLUSIONS

It has been shown that, if it is permissible to ignore harmonics generated by non-linear elements within a system, frequency-response analysis leads both to qualitative and quantitative results concerning the behaviour of a non-linear system. The accuracy that can be achieved has been illustrated by examples of a control system in which the only non-linearity is an element of the on-off type in the forward loop. In all cases the results agree to within about 10% of accurate solutions. The "on-off" element is the extreme example of a soft-spring characteristic, and hence it is reasonable to assume that even better accuracy would be obtained with other soft-spring characteristics.

However, there is, in principle, no limitation to the form of the non-linear elements involved, the only requirement being that there is sufficient attenuation of harmonics in the various loops of the system. Nor is frequency-response analysis confined to control systems; many other non-linear systems can be analysed in this way, but a block schematic or some other guide is necessary in order to determine whether it is reasonable to ignore the presence of harmonics.

The argument used in Section 3.3 for predicting the stability of steady oscillations is not entirely sound for two reasons, quite

apart from the inherent approximation of the frequency-response analysis. First, there is the possibility of other modes of oscillation affecting the mode considered. An inspection of the Nyquist diagram may show what to expect in such cases, and the example of Section 3.4 illustrates this point. Secondly, the criterion is based on the assumption that near a steady-state solution the damping is large compared with its derivatives. The analysis given in Section 9 shows that this assumption cannot be sustained, and that a different, but less convenient, criterion of stability should be used. Experience suggests, however, that to distinguish between the two criteria is more of academic than practical importance, since no system has yet been found for which they give different answers.

Transient solutions can be obtained for non-linear systems which are governed by second-order differential equations, i.e. of one degree of freedom. It is not practicable to obtain the full transient for higher-order systems, owing to the intermodulation terms which would be generated by the non-linear elements. A further restriction is that the required solution should be oscillatory in form. Obtaining the transient solution involves finding a good approximate solution to the non-linear equations for frequency and damping. From an examination of these equations much can be learnt concerning the nature of the transient, but it may not be easy to obtain simple analytical solutions to the equations for examples other than those given in Section 5.

(7) ACKNOWLEDGMENTS

The author wishes to thank Mr. R. H. Macmillan for much useful advice and detailed discussion throughout the preparation of the paper. The author also wishes to thank the Department of Scientific and Industrial Research for the receipt of a maintenance grant during the same period.

(8) REFERENCES

(1) TUSTIN, A.: "The Effects of Backlash and of Speed-Dependent Friction on the Stability of Closed-Cycle Control Systems," *Journal I.E.E.*, 1947, **94**, Part IIA, p. 143.

(2) KOCHENBURGER, R. J.: "A Frequency Response Method for Analysing and Synthesizing Contactor Servomechanisms," *Transactions of the American I.E.E.*, 1950, **69**, Part I, p. 270.

(3) JOHNSON, E. C.: "Sinusoidal Analysis of Feedback-Control Systems containing Non-Linear Elements," *ibid.*, **71**, Part II, p. 169.

(4) GRIEF, H. D.: "Describing Function Method of Servomechanism Analysis applied to most Commonly Encountered Non-Linearities," *ibid.*, 1953, **72**, Part II, p. 243.

(5) NICHOLS, B. N.: "Backlash in a Velocity Lag Servomechanism," *ibid.*, 1953, **72**, Part II, p. 462.

(6) SLEMON, G. R.: "A Method of Approximate Steady-State Analysis for Non-Linear Networks," *Proceedings I.E.E.*, Paper No. 1530, September, 1953 (**100**, Part I, p. 275).

(7) JACOBSEN, L. S.: "Steady Forced Vibration as influenced by Damping," *Transactions of the American Society of Mechanical Engineers, Applied Mechanics*, 1930, **53**, p. 169.

(8) KLOTTER, K.: "Steady State Vibrations in Systems having Arbitrary Restoring and Arbitrary Damping Forces," Proceedings of the Symposium on Non-Linear Circuit Analysis, Polytechnic Institute of Brooklyn, 1953, p. 234.

(9) STOKER, J. J.: "Non-Linear Vibrations" (Interscience Publishers, 1950).

(10) KRYLOV, N., and BOGOLUIBOV, N.: "Introduction to Non-Linear Mechanics," Free translation by S. Lefschetz (Princeton University Press, 1943).

(9) APPENDIX

The purpose of the Appendix is to establish the necessary modifications which must be made to the familiar j notation of linear-system analysis, so that this notation may also be used for non-linear systems.

Consider an equation of the form

$$\mathrm{P(D)}x = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

where D is the operator d/dt, and P(D) is a polynomial in D. The solution is obtained by the substitution

$$x = A\varepsilon^{(\alpha+j\omega)t}$$

and eqn. (42) becomes

$$\mathrm{P(D)}A\varepsilon^{(\alpha+j\omega)t} = 0$$

or, by a well-known operational theorem,

$$A\varepsilon^{(\alpha+j\omega)t}\mathrm{P(D} + \alpha + j\omega) = 0$$

or, since $A \neq 0$,

$$\mathrm{P(D} + \alpha + j\omega) = 0 \quad . \quad . \quad . \quad . \quad . \quad (43)$$

Finally, since D is operating only on constants in eqn. (43), D may be made equal to zero, and the solution of eqn. (42) becomes the solution of

$$\mathrm{P}(\alpha + j\omega) = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

The damping and frequency of the various modes of oscillation are $-\alpha$ and ω, respectively.

Suppose now that α and ω are not constant, then for convenience we write

$$\alpha + j\omega = z = \mathrm{z}(t)$$

In order to carry out the substitution used in the case of constant values of α and ω, use is made of the following theorem, which states that

$$\mathrm{P(D)}\varepsilon^{\int z dt}\phi(t) \equiv \varepsilon^{\int z dt}\mathrm{P(D} + z)\phi(t) \quad . \quad . \quad . \quad (45)$$

The interpretation of $\mathrm{P(D} + z)$ requires care, for on expansion it will contain powers of z, the derivatives of z and the operator D. In particular $\mathrm{P(D} + z)$ is the sum of terms of the form $(\mathrm{D} + z)^n$, and such terms must be multiplied out, maintaining the correct order of z and D.

Thus $(\mathrm{D} + z)^2 = (\mathrm{D} + z)(\mathrm{D} + z) = \mathrm{D}^2 + \mathrm{D}z + z\mathrm{D} + z^2$

Also, since $\mathrm{D}(z\phi) = (\dot{z} + z\mathrm{D})\phi$

we may write $\mathrm{D}z = \dot{z} + z\mathrm{D}$

Thus, finally, $(\mathrm{D} + z)^2 = \mathrm{D}^2 + 2z\mathrm{D} + z^2 + \dot{z}$

Similarly

$$(\mathrm{D} + z)^3 = \mathrm{D}^3 + 3z\mathrm{D}^2 + 3(z^2 + \dot{z})\mathrm{D} + z^3 + 3z\dot{z} + \ddot{z}$$

With the above interpretation of $\mathrm{P(D} + z)$ established, the theorem is simply proved by induction for the case where $\mathrm{P(D)} \equiv \mathrm{D}^n$, and then the extension to the case where P(D) is a polynomial in D follows at once.

The theorem is now applied to the equation for the free oscillations of the control system described in Section 3. We assume that oscillations of only one frequency, $\omega = \omega(t)$, occur, and that the amplitude of such an oscillation is $a = \mathrm{a}(t)$. If the harmonics introduced by the non-linear element are ignored the error signal in the system may be represented by

$$e = \varepsilon^{\int(-\mu+j\omega)dt} \quad . \quad . \quad . \quad . \quad . \quad (46)$$

and the amplitude of the error by

$$a = \varepsilon^{-\int \mu dt} \quad . \quad . \quad . \quad . \quad . \quad . \quad (47)$$

The gain of the system is $\mathrm{n}(a)$ as defined by eqns. (5), (6) and (8).

In this notation the equation of the system becomes

$$[\mathrm{R(D)n}(a) + \mathrm{S(D)}]\varepsilon^{\int(-\mu+j\omega)dt} = 0 \quad . \quad . \quad (48)$$

where R(D) and S(D) are polynomials in D, and the transfer operator of the linear filter is

$$\mathrm{G(D)} \equiv \mathrm{R(D)/S(D)} \quad . \quad . \quad . \quad . \quad . \quad (49)$$

By use of the theorem already discussed, eqn. (48) becomes

$$\mathrm{R}(\mathrm{D} - \mu + j\omega)\mathrm{n(a)} + \mathrm{S}(\mathrm{D} - \mu + j\omega) = 0 \quad . \quad (50)$$

Further simplification is not possible in general; the operator D must be retained since n(a) is not constant, and hence μ and ω are not constant during a transient.

However, if only the steady-state solution is looked for, we have

$$a = a_s = \text{constant},$$
$$\omega = \omega_s = \text{constant},$$
and $$\mu = 0.$$

All the variables in eqn. (50) are thus constant, and the steady-state solution becomes

$$\mathrm{G}(j\omega_s)\mathrm{n}(a_s) + 1 = 0$$

an equation derived in the text.

Another solution which may be obtained is that near a steady-state solution. This allows the examination of the stability of the steady-state solution.

Suppose that near a steady-state solution

$$-\mu + j\omega = j\omega_s + \Delta z \quad . \quad . \quad . \quad . \quad . \quad (51)$$

and also that

$$a = a_s + \Delta a \quad . \quad . \quad . \quad . \quad . \quad (52)$$

Δz is complex and assumed to be of the form

$$\Delta z = \xi \varepsilon^{kt} \quad . \quad . \quad . \quad . \quad . \quad . \quad (53)$$

where ξ is a complex constant and k is a real constant. Thus it is assumed that the damping and frequency converge to or diverge from their steady-state values exponentially.

The behaviour of Δa with time may now be written down.

From eqn. (47) $$\dot{a}/a = -\mu$$

i.e. $$\frac{d}{dt}(\Delta a) = -a\mu$$
$$= -a_s\mu$$

to the order of accuracy required. But since, from eqn. (51) $-\mu = \mathscr{R}(\Delta z)$, we have the result that, if Δa is small,

$$\Delta a = \text{constant} \times \varepsilon^{kt}$$

It is clear that the stability of the steady-state solution depends only on the sign of k; an equation is now derived from which the value of k may be determined.

From eqn. (53) $k = \dot{\mu}/\mu$,

i.e. from eqn. (47) $k = -a\dot{\mu}/\dot{a} = -a(d\mu)(da)$

$$= a_s(\mathscr{R}\Delta z)/\Delta a \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

to the order of accuracy required.

From eqn. (50) we have

$$\mathrm{R}(\mathrm{D} + j\omega_s + \Delta z)\mathrm{n}(a_s + \Delta a) + \mathrm{S}(\mathrm{D} + j\omega_s + \Delta z) = 0$$
$$. \quad . \quad . \quad (55)$$

and it is possible to expand the left-hand side of this equation in powers of Δa and Δz. By the use of Taylor's theorem

$$\mathrm{n}(a_s + \Delta a) = \mathrm{n}(a_s) + \Delta a\mathrm{n}'(a_s) + [\text{Terms of higher order in } \Delta a]$$
$$. \quad . \quad . \quad . \quad (56)$$

If Δz is of the form given in eqn. (53), there is a theorem which may be proved by the use of eqn. (45) that states

$$\mathrm{P}(\mathrm{D} + \Delta z) = \mathrm{P(D)} + \Delta z\left[\frac{\mathrm{P}(\mathrm{D} + k) - \mathrm{P(D)}}{k}\right]$$
$$+ [\text{Terms of higher order in } \Delta z] \quad . \quad (57)$$

Using the expansions of eqns. (56) and (57) on eqn. (55) and substituting the expression obtained for $\Delta z/\Delta a$ into eqn. (54) leads to the result that

$$k = -\mathscr{R}\left\{[a_s\mathrm{n}'(a_s)R(j\omega s + k)]\Big/\left[\mathrm{n}(a_s)\frac{\mathrm{R}(j\omega_s + k) - \mathrm{R}(j\omega_s)}{k} + \frac{\mathrm{S}(j\omega_s + k) - \mathrm{S}(j\omega_s)}{k}\right]\right\} \quad . \quad (58)$$

The value of k obtained from this equation indicates the stability and quantitative behaviour of a transient near a steady-state solution.

In the text the stability was determined by considering the solution of eqn. (50) or (55) with D made equal to zero. Such an assumption is not justified, for it leads to appreciably different results from the more rigorous analysis used above. Thus with D made equal to zero in eqn. (55) we find

$$\Delta a = \text{constant} \times \varepsilon^{k't}$$

where

$$k' = -\mathscr{R}\left\{[a_s\mathrm{n}'(a_s)\mathrm{R}(j\omega_s)]/[\mathrm{n}(a_s)\mathrm{R}'(j\omega_s) + \mathrm{S}'(j\omega_s)]\right\} \quad . \quad (59)$$

Only if $|k| \ll \omega_s$ do the denominators of eqns. (58) and (59) become identical, and in general k and k' differ appreciably. In example (d) of Section 3.5 with $T_1 = T_2 = T$, and hence $\omega_s T = 1$, we find $kT = -1/3$ approximately, whereas $k'T = -1/5$.

However, in all practical examples it appears that k and k' are of the same sign, and for this reason the simpler method of determining the stability used in the text may be adequate.

621.315.61:621.3.011.5

Monograph No. 127 M
Apr. 1955

BREAKDOWN OF SOLID DIELECTRICS IN DIVERGENT FIELDS

By J. H. MASON, B.Sc., Associate Member.

(*The paper was first received* 3*rd November*, 1954, *and in revised form* 11*th January*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in April*, 1955.)

SUMMARY

Breakdown channels are shown to propagate from steel needle-point electrodes embedded in polythene and polyisobutylene when the maximum effective stress at the end of the point reaches the intrinsic electric strength of the material at the test temperature. The average breakdown stress shows no significant variation with the radius of curvature of the point (for radii of 1–20 micron), but, at room temperature, about 25% greater stress is required with a negative than with a positive point. These results are explained by the effect of enhanced conductivity and space-charge accumulation increasing the effective point radius to some 25 microns for positive and about 45 microns for negative points. The effect of polarity decreases with increasing temperature and is negligible at 100°C.

The decrease of the average stress required for breakdown with increasing electrode separation is explained, and the same concept is used to predict the variation of the industrial electric strength of materials with specimen thickness.

Factors affecting the industrial electric strength of materials are discussed and illustrated by tests on polythene, polystyrene and cellulose acetate, using British Standard electrodes in air and in clean and contaminated transformer oil.

LIST OF PRINCIPAL SYMBOLS

E_{max} = Calculated maximum stress at point electrode.
Z, ρ, ξ = Co-ordinates of hyperboloids.
c_2 = Function of Z and ξ.
$\hat{\xi}$ = ξ_0 at tip of hyperboloid.
t = Distance between point and plane (Fig. 2); in Sections 5.2 and 9.2 it denotes thickness of specimen.
R = Radius of curvature of point.
V_0 = Potential of point electrode.
E = Stress at points on Z axis.
E_{av} = Average stress between point and plane.
E_1 = Stress at tip of spheroidal boss.
c = Length of boss (Fig. 2).
t' = $(t + c)$ in Fig. 2.
E_I = Intrinsic electric strength of dielectric.
ΔW = Energy difference between conduction levels, and mid-point of imperfection levels.
k = Boltzmann's constant.
T_0 = Absolute temperature of electrons in solid.
σ = Conductivity.
σ_0 = Conductivity in absence of applied field.
W = Energy difference of imperfection levels below conduction zone.
W/A = Discharge energy per unit area (Section 5.2).
E_M = Electric strength of a length l of medium.
ϵ_M; ϵ_D = Permittivity of medium and specimen respectively.
V_i = Discharge inception voltage.
V = Voltage applied to specimen.
t_T = Distance between point and plane at temperature T.
γ = Coefficient of cubical expansion.

Correspondence on Monographs is invited for consideration with a view to publication.
The paper is based on Report Ref. L/T310 of the British Electrical and Allied Industries Research Association.

(1) INTRODUCTION

Investigations of the breakdown of polythene by internal discharges[1,2] indicate two distinct mechanisms leading to failure. The first mechanism occurs at stresses just exceeding the discharge inception value, when discharges cause, initially, slow erosion of the dielectric, at a rate corresponding either to thermal degradation by the energy of the discharge or to disruption of the carbon–hydrogen bonds by bombardment. The rate of erosion increases rapidly with voltage, and the discharges concentrate to form deep pits. The energy liberated by each discharge increases with its length, and the erosion propagates with increasing rapidity until the pits attain a critical length, when narrow semi-carbonized channels develop at their ends and often precipitate complete breakdown. It can be shown that there is intense field-concentration at the ends of the pits, and it is believed that the ultimate breakdown channels are propagated when the stress exceeds the intrinsic electric strength over some minimum distance. If the applied stress is sufficient these channels may propagate immediately, without preliminary erosion, and this second mechanism occurs in short-time industrial electric-strength tests.[3] Sometimes excessive carbonization short-circuits the discharges in the channels, leaving the latter as pointed conducting tracks, and it is thus of practical interest to determine whether breakdown channels propagate under the same divergent field conditions in the absence, as in the presence, of discharges.

The propagation of breakdown channels from needle-point electrodes embedded in polythene to adjacent plane electrodes has therefore been investigated. Under such conditions, in the absence of discharges, breakdown channels propagate from the point when the effective electric stress in the vicinity reaches the intrinsic electric strength of the dielectric. The formation of space-charge around the point, caused presumably by stress-enhanced conduction, is believed to decrease the effective stress at the point, which may therefore be much less than the value calculated from the geometry of the system. Thus the calculated stress at the point, denoted by E_{max}, does not represent a stress actually existing in the material, because this is limited by conduction to a maximum value equal to the intrinsic electric strength.

This report describes and discusses the dependence of the divergent-field electric strength on (*a*) the radius of curvature and polarity of the point electrode, (*b*) the electrode separation, and (*c*) the ambient temperature. The effects of repeated pulses of the same polarity, of successive pulses of opposite polarity, and of alternating voltage are also considered.

(2) TEST PROCEDURE

(2.1) Experimental Technique

The point electrode was mounted in a special holder, shown in Fig. 1, which was so designed that the point could be inserted, perpendicularly and without rotation, to a predetermined depth into the dielectric.

Nickel-plated-steel sewing needles were used for the investigation. Some tests were made with the needles as manufactured,

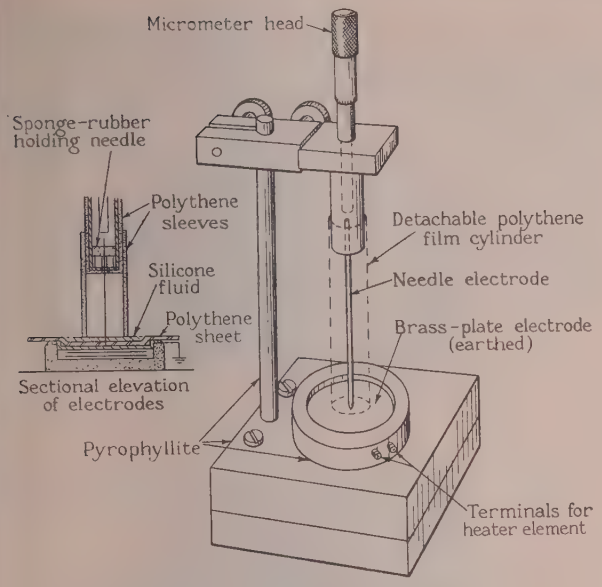


Fig. 1.—Electrodes for divergent-field breakdown tests.

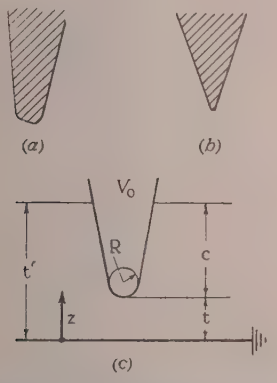


Fig. 2.—Typical point electrodes.
(*a*) 25 micron radius of curvature.
(*b*) 1·5-micron radius of curvature.
(*c*) Electrode nomenclature.

the points having radii of curvature between 25 and 30 microns, as shown in Fig. 2(*a*). For the majority of tests the points were ground to radii of about 2 or 5 microns, as shown in Fig. 2(*b*). It is recognized that the work function of exposed steel may differ from that of the original nickel-plated-steel needles, but the effect, if any, is probably small.* Tests will be made later to determine the influence of the electrode work function on the electric strength in divergent-field conditions.

The radius of curvature of the points was measured to an accuracy of within $\pm\frac{1}{2}$ micron or $\pm 10\%$, whichever was greater. In some tests with polythene the needle was withdrawn from the dielectric after applying insufficient voltage to cause breakdown. Microscopic examination showed that the point was then undamaged. If the polythene from which the point was withdrawn was sectioned and examined under a polarizing microscope, there was evidence of some strain around the point, but there were no cracks or bubbles in the polythene, neither was there any evidence to suggest that the point had not been in intimate contact with the polythene.

Preliminary tests were made with films of polyisobutylene, of about 1 mm thickness, cast on brass plates, from solution in petroleum ether (boiling point, 100–120° C). The needles were readily pressed into this material, but because of its plasticity and low mechanical strength, the samples could not afterwards be removed and sectioned for microscopic examination.

The majority of tests were therefore made using sheets of polythene grade 7 (molecular weight, about 17 000) of about 0·6–1·6 mm thickness. These sheets were melted on the special plate-electrode shown in Fig. 1, using a 100-watt heater element incorporated below the electrode. The needle point was inserted into the molten polythene, which was then cooled to the required test temperature (indicated by a copper-constantan thermojunction inserted between the plane electrode and the polythene sheet, at the side farthest from the point electrode). The electrodes were mounted in an oven which was adjusted to the same temperature as the specimen.

A small quantity of silicone fluid (viscosity, about 1 000 centistokes) was poured on the polythene surface, around the needle-point, to prevent surface discharges which might cause spurious breakdown. A cylinder of polythene film was also placed around the needle support-column to raise the flashover voltage, which otherwise limited the maximum test voltage to about 30 kV peak. It is believed that field distortion in the air would not appreciably affect the stress at the electrode point.

Alternating, surge or successive unidirectional pulse voltages were applied to the specimen by means of a high-voltage switch, synchronized with a rotating-drum camera and cathode-ray-oscillograph trapping system. The time of voltage application was adjustable from about 1/25 sec upwards. A fraction of the applied voltage was recorded oscillographically, as shown in Fig. 3, using a calibrated resistance voltage divider, so that the maximum applied voltage and the voltage at which breakdown actually occurred could be estimated.

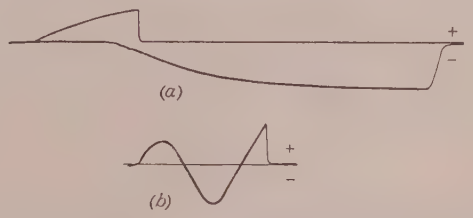


Fig. 3.—Oscillograph records showing that, in polythene at 20° C, divergent-field breakdown occurs with a lower positive than negative potential on the point.
(*a*) Breakdown with 32 kV surge (positive) after withstanding 32 kV (negative) and 52 kV (negative) surges (R = 22 microns; t = 470 microns).
(*b*) Breakdown at 40 kV (positive) peak after one half-cycle of 45 kV (negative) peak (R = 3 microns; t = 650 microns).

(2.2) Evaluation of Electric Stresses

The stress at the end of the point electrode was calculated on the assumption that it can be represented by a conducting hyperboloid of the form

$$\frac{Z^2}{c_2^2\xi_0^2} - \frac{\rho^2}{c_2^2(1-\xi_0^2)} = 1 \quad . \quad . \quad . \quad . \quad (1)$$

If this hyperboloid is at a potential V_0 and a distance t from an earthed plane electrode, as shown in Fig. 2(*c*), the stress at points on the Z-axis is given[1] by

$$E = 2V_0/c_2(1-\xi)^2 \log(1+\xi_0)/(1-\xi_0) \quad . \quad . \quad (2)$$

* The thermionic work function of steel is 4·5 eV and that of nickel is 4·6 eV. The photo-electric work functions are respectively 4·6 and 5·0 eV.[5] Recent papers[6,7] indicate that such increase in the photo-electric work function of the electrodes causes about a 10% increase in the electric strength of liquids, but similar variation with solid dielectrics has not been confirmed.[8]

On the Z-axis $Z = c_2\xi$ and at the tip $Z = t$, $\xi = \xi_0$. The radius of curvature at the tip is given by $R = t(1 - \xi_0^2)/\xi_0^2$, so that the stress at a distance $(t - z)$ from the point can be written

$$E = 2V_0 p/\log q \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where $p = t^{1/2}(t + R)^{1/2}/(t^2 - z^2 + tR)$

and $q = [2t + R + 2t^{1/2}(t + R)^{1/2}]/R$

At the tip, where $z = t$,

$$E_{max} = 2V_0 p_1/\log q = 2E_{av} t p_1/\log q \quad . \quad . \quad (4)$$

where $p_1 = (1 + R/t)^{1/2}/R$

and $E_{av} = V_0/t$ is the average electric stress between the point and plane.

If $t > 10R$, the stress at the point can be written

$$E_{max} \simeq 2E_{av} t/R \log (1 + 4t/R) \quad . \quad . \quad . \quad (5)$$

The value of E_{max} is directly affected by errors in determining V_0 and R and to a smaller extent by error in t. Fortunately V_0 and R are directly measurable to an accuracy of about $\pm 10\%$, but t was calculated as the difference between the total dielectric thickness and the depth of penetration of the needle point; some error may therefore enter because of the large coefficient of thermal expansion of polythene.* The stress at the needle tip was also calculated from an expression for the stress E_1 at the end of a conducting spheroidal boss, which projects from a conducting plate into a semi-infinite dielectric,[1]

i.e. $$E_1 = 2E_0 c/R \, [\log (4c/R) - 2] \quad . \quad . \quad . \quad . \quad (6)$$

where, as in Fig. 2(*c*), $E_0 = V_0/t'$ is the initial stress in the dielectric, c is the length of the pit and t' the total thickness of dielectric.

The calculated values of E_1, which are independent of t, were some 5–15% lower than the values of E_{max} calculated from eqn. (4). This difference is not significant compared with the effect of varying the radius of curvature and polarity of the needle point, so that values of E_{max} only are quoted in the paper.

(3) RESULTS

(3.1) Tests with Polyisobutylene

The initial tests with polyisobutylene, which are summarized in Table 1, indicate that

(*a*) For a given electrode configuration the a.c. electric strength is lower than the d.c. value.

(*b*) The radius of curvature of the needle point may not appreciably affect the average electric strength.

(*c*) Breakdown occurs only when the calculated maximum electric stress at the tip considerably exceeds the intrinsic electric strength of polyisobutylene (i.e. 2 MV/cm at 20° C).[9]

* See Appendix 9.1 for details of a correction which has been applied.

(3.2) Tests with Polythene

The photomicrographs in Fig. 4 show the form of initial channels from point electrodes, and complete breakdown channels with associated partial breakdown channels. It is interesting to

(*a*) (*b*)

(*c*)

Fig. 4.—Breakdown of polythene in the absence of discharges in a divergent field between a needle point and a plane electrode.

(*a*) Partial breakdown channel after surges of 52 kV (negative) and 60 kV (negative) peak ($R = 1{\cdot}5$ microns; $t = 600$ microns; $E_{av} = -1{\cdot}0$ MV/cm).
(*b*) Partial breakdown channels after 5 cycles of 23 kV peak ($R = 2$ microns; $t = 420$ microns; $E_{av} = 0{\cdot}45$ MV/cm).
(*c*) Breakdown at 40 kV (positive) peak after one cycle at 50 kV (a.c.) peak ($R = 32$ microns; $t = 590$ microns; $E_{av} = 1{\cdot}1$ MV/cm).

Table 1

ELECTRIC STRENGTH OF POLYISOBUTYLENE IN DIVERGENT FIELDS AT 20° C

Number of samples tested	Radius of tip, R	Space between tip and plate, t	Type of voltage applied	E_{av}, peak	E_{max}, peak
	microns	cm		MV/cm	MV/cm
4	4 ± 1	0·035	About two cycles of alternating current	$0{\cdot}25 \pm 0{\cdot}05$	$7{\cdot}7 \pm 1{\cdot}5$
1	15	0·065		0·26	4·5
4	5 ± 2	0·03	D.C.* surges of 1/10 sec duration	$0{\cdot}5 \pm 0{\cdot}15$	12 ± 2
1	17	0·05		0·32	4·0

* The polarity of the d.c. surges was not recorded.

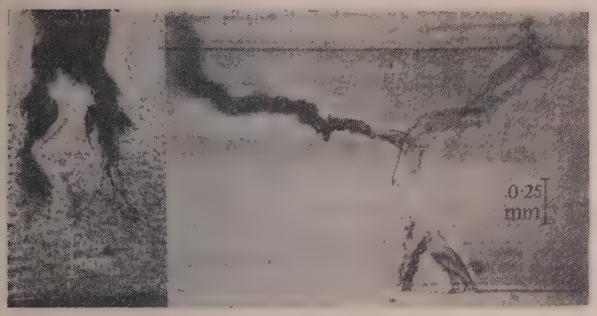


Fig. 5.—Breakdown channels in polythene caused by discharges.

(*a*) After 74 hours at 150kc/s at $2V_i$ (100kV/cm). Discharges in a void of 0·2mm depth, in a disc 0·6mm thick.
(*b*) After about 5 cycles at more than 500kV/cm; with a 1·25mm thick sample between A.S.T.M. electrodes under transformer oil.

note the similarity between these channels and others, shown in Fig. 5, which were caused by internal or surface discharges.

The curves in Fig. 6 show the variation of the average divergent-field electric strength (E_{av}) with temperature, for the range 20–100°C, using points of between 1·5 and 7·5-micron radius, and electrodes spaced between 250 and 800 microns apart. For comparison the variation of the intrinsic electric strength with temperature[10] is shown in Fig. 6(*d*). With points of any given radius the variation of E_{max} with temperature is similar to that shown in Fig. 6 for E_{av}.

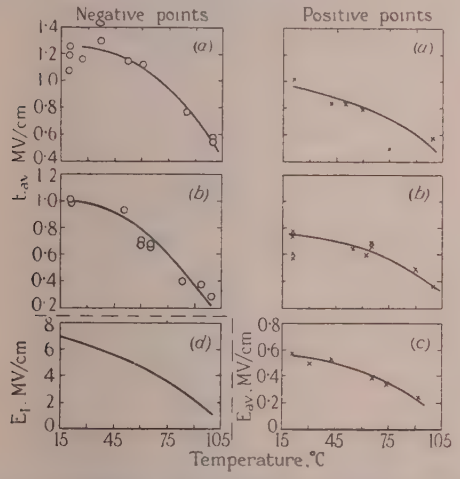


Fig. 6.—The variation of the intrinsic and average divergent-field electric strengths of polythene with temperature. Results for points of between 1·5 and 7·5 microns radius.

The electrode separation was as follows:
(*a*) 250–350 microns.
(*b*) 450–600 microns.
(*c*) 750–800 microns.
(*d*) Variation of intrinsic electric strength with temperature.

Despite considerable scatter the curves in Fig. 6 indicate that

(*a*) E_{av} and the intrinsic strength vary with temperature in a similar manner.

(*b*) At 20°C, the breakdown stress for a given electrode configuration is some 25% greater if the point is negative, than if it is positive. This polarity effect decreases with increasing temperature and is negligible above 60°C when the electrodes are 500 microns apart.

(*c*) E_{av} decreases with increasing separation between point and plane.

Tests were also made with larger points of up to 50-micron radius. Then, with the same temperature range (20–100°C), and electrodes spaced between 250 and 800 microns apart, it is found that E_{av} does not vary significantly with the size of the point, unless the radius R exceeds about 10 microns, as shown in Fig. 7. With larger points E_{av} increases rapidly with R, such

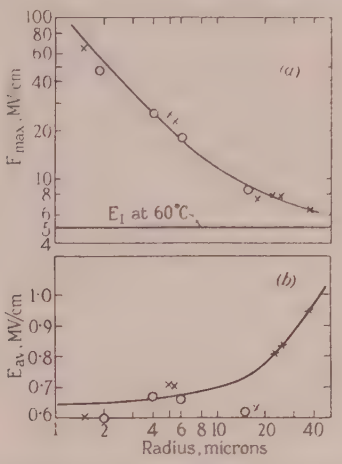


Fig. 7.—Variation of divergent-field electric strength with radius of point at 60°C.

Electrode separation, 500 microns.
× × × Positive points.
○ ○ ○ Negative points.

that at breakdown the maximum stress at the point [E_{max}, calculated from eqn. (4)] always exceeds the intrinsic strength, as shown in Fig. 7(*a*).

If a surge of one polarity, which is insufficient to cause breakdown, is followed* by a surge of opposite polarity, failure occurs at much lower electric stress than normal for this polarity.

Representative results given in Table 2 show that

(*a*) After a positive surge (of 50% of the positive breakdown voltage), a subsequent negative surge causes failure at only 50% of the normal negative breakdown stress.

(*b*) After negative surges of about 40 and 80% of the negative breakdown voltage, subsequent positive surges cause failure at about 70% of the normal positive breakdown stress.

With alternating voltage, breakdown occurs always when the point is positive, often after one complete voltage cycle, as shown in Fig. 3(*b*). If breakdown occurs in the second positive half-cycle, the breakdown strength is about 25% lower than with a single positive pulse, but failure can occur at a lower stress after a large number of cycles, as shown by Table 2.

(4) DISCUSSION

For convenience the results of principal interest are discussed under separate headings.

(4.1) Similarity between the Temperature Characteristics of the Intrinsic and the Divergent-Field Electric Strengths

The similar variation with temperature of the intrinsic and divergent-field electric strengths of polythene, shown in Fig. 6, seems to confirm that both types of failure are controlled by the same basic mechanism.

The criterion for intrinsic breakdown in solids[10] is that free electrons in the conduction band gain more energy from the

* In the present tests there was an interval of about 5 min between tests with opposite polarity.

Table 2

Effect of Preceding Test Conditions on Breakdown Strength

Radius of tip, R	Space between tip and plate, t	Test conditions preceding breakdown	Breakdown strength	
			E_{av}, peak	E_{max}, peak
microns	cm		MV/cm	MV/cm
1·5	0·060	Breakdown with a single negative surge	−1·0	−108
1·5	0·059	Surge of $E_{av} = +0{\cdot}32$ MV/cm followed by negative surge	−0·51	−55
2·5	0·058	Breakdown with a single positive surge (a)	+0·76	+54
		(b)	+0·60	+43
3·0	0·058	Breakdown in positive half-cycle after negative half-cycle of $E_{av} = -0{\cdot}40$ MV/cm	+0·46	+27
2·0	0·059	Breakdown after 55 cycles of $E_{av} = 0{\cdot}29$ MV/cm	0·29	24
2·0	0·059	Breakdown after 1 750 cycles of $E_{av} = 0{\cdot}27$ MV/cm	0·27	23
4·0	0·067	Breakdown after 3 cycles of $E_{av} = 0{\cdot}60$ MV/cm	+0·55	+28
8·0	0·078	Breakdown at +0·45 MV/cm after previous surge of −0·8 MV/cm	+0·45	+14
32	0·059	Breakdown after 2 cycles of $E_{av} = 0{\cdot}85$ MV/cm	+0·66	+5·8

applied field than they lose in collision with the lattice, or by interaction with electrons trapped in imperfection levels. It can be shown that the intrinsic strength E_I is given by

$$\log E_I = \text{constant} + \Delta W/2kT_0 \quad . \quad . \quad . \quad (7)$$

where T_0 is the electron temperature, k is Boltzmann's constant, and ΔW is the energy difference between the conduction level and the mid-point of the imperfection levels.

The same concept shows that the conductivity, σ, varies with field strength E, as follows

$$\log (\sigma/\sigma_0) \propto (W/\Delta W)(E^2/E_I^2) \quad . \quad . \quad . \quad (8)$$

where σ_0 is the conductivity in the absence of applied field, W is the potential barrier of trapped electrons, and ΔW and E_I are the same as in the preceding paragraph. Thus it is to be expected that the conductivity of the dielectric in the highly-stressed region near a point electrode will be greater than in the bulk of the material. The effective stress near the point is consequently reduced, as shown mathematically by Flynn.[24]

The greatest reduction in effective stress is expected when the ratio R/t (radius of point electrode)/(electrode separation), is small, as shown in Fig. 7 of Flynn's paper. If breakdown channels propagate only when the effective stress near the point attains the intrinsic strength E_I, it follows that the calculated stress at the point, E_{max}, corresponding to the breakdown voltage should decrease with increasing R/t. This is confirmed by the results in Fig. 7(*a*) of this paper, which show that at 60°C, with $t = 500\mu$, E_{max} decreases from 65 MV/cm with $R = 1{\cdot}5\mu$ to 6 MV/cm when $R = 37\mu$. It is also found that E_{max}/E_I varies with R/t in a similar manner at 20, 60 and 100°C, showing that the effects of field-dependent conductivity are similar throughout this temperature range.

(4.2) Variation of the Divergent-Field Electric Strength with the Polarity of the Point Electrode

The greater temperature dependence of E_{av} with negative than with positive points, shown in Fig. 6, indicates that field-dependent conductivity is not the only factor controlling the divergent-field electric strength.

The polarity effect at room temperature can probably be explained as follows: when a high electric stress is applied, electrons are emitted from negative points and penetrate the dielectric. In regions of lower stress the electrons lose energy, in collisions and become trapped, forming a space-charge round the point. The electrons penetrate the dielectric furthest along the axis, where the potential gradient is greatest, and to a lesser extent in other directions. The space charge may increase the effective radius and length of the point, and decrease the maximum stress at the point, as shown diagrammatically in Fig. 8,

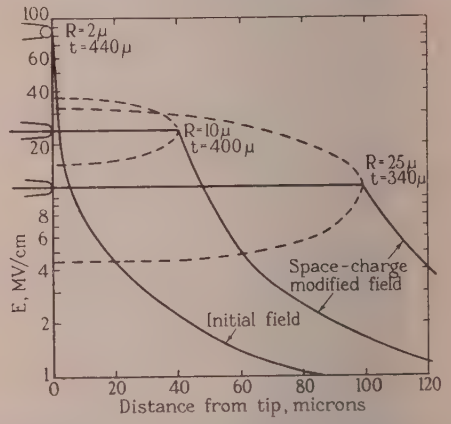


Fig. 8.—Probable effect of space charge on the field near a point electrode of 2-micron radius. $V_0 = 60$ kV (negative) peak.
$t = 440$ microns.
$E = 2V_0 p/\log q$ (eqn. 3).

where it is arbitrarily assumed (for the purpose of illustration only) that the effective length of the point propagates four times as rapidly as the increase of effective radius of curvature.

With a positive point electrons will move towards the point, leaving a positive space-charge round the point. The number of free electrons in polythene is limited, however, especially at low temperatures, so that the space charge around a positive point may not decrease the effective stress so much as with a negative point. Breakdown can thus occur at a lower voltage with a positive than with a negative point.

At higher temperatures the polarity of the point has less influence on the effective electric stress near the point, and consequently little or no influence on the divergent-field strength, as shown in Figs. 6 and 7. This is because

(*a*) At the lower stress required for intrinsic breakdown there is less field emission from negative points than at room temperature.

(*b*) The number of free electrons in the dielectric is greater so that there is greater space-charge near positive points.

Evidence for the influence of space charge is provided also by

the results in Table 2, which show that the application of successive surges of opposite polarity causes considerable reduction in the divergent-field strength. Evidently space charge persists after the application of an initial surge and increases the effective stress when a surge of opposite polarity is applied subsequently.

(4.3) Decrease of Average Stress required for Breakdown with increasing Electrode Separation

The points in Fig. 9 show the variation of the average breakdown stress at 20°C with gap length t for points having a

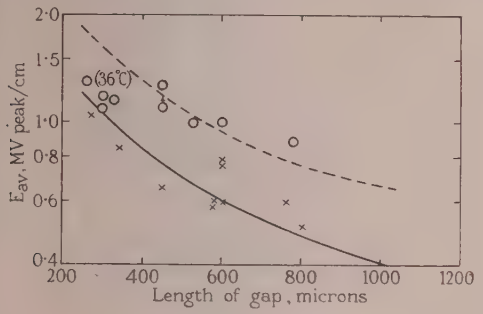


Fig. 9.—Relation between breakdown strength, E_{av}, and length of gap (t) at 20°C.

○○ Negative points, ×× Positive points } Measured with R = 2–10 microns.
———— calculated for R = 25 microns
— — — — calculated for R = 45 microns } E_{max} = 6·5 MV/cm.

radius of curvature of between 2 and 10 microns, while the curves show calculated values of E_{av} required to give values of E_{max} equal to 6·5 MV/cm at points of 45- and 25-micron radius. There are too few experimental values to establish definite correlation, but it can be inferred from Fig. 9 that with positive points the effective radius is increased to about 25 microns. With negative points the effective radius may be increased to some 45 microns when the electrode separation exceeds 500 microns.

The industrial electric strength* of dielectric materials also decreases with specimen thickness, as shown by the points in Fig. 11, and many empirical formulae relating breakdown voltage and thickness have been derived.[11] The subsequent discussion indicates that the variation with thickness can be successfully predicted by a formula based on physical concepts.†

(5) FACTORS AFFECTING THE INDUSTRIAL ELECTRIC STRENGTH OF INSULATION

(*a*) Recent work[3] shows that in industrial tests breakdown occurs when the stress in the dielectric is sufficient for discharges in the medium to penetrate the surface and propagate through the specimen. The present paper shows that this could occur if the effective stress at the end of the discharge channel exceeded the intrinsic electric strength of the material.

(*b*) Surface discharges are believed to cause considerable local temperature rise at the dielectric surface,[9] so that breakdown may occur when the high temperature intrinsic strength is exceeded, irrespective of the ambient temperature.

(*c*) Electro-photographs and calculations based on the discharge magnitude show that individual internal discharges in polythene[1] affect an area of radius about 0·1–0·5 mm (4–20 mils). External discharges in air probably affect a greater area, but in oil discharges are more concentrated,[12] so that it is reasonable to assume that in industrial tests discharges impinge on the dielectric surface with an effective radius of curvature between 1 and 20 mils.

The local stress at which discharge channels propagate through a given thickness of material can be calculated from the industrial electric strength E_{av} if eqn. (4) is assumed to give the electric stress E_{max} at the point of impact of a discharge of radius R, on dielectric of thickness t. Table 3 shows values of E_{max}

Table 3

CALCULATED VALUES OF STRESS E_{max} AT THE END OF DISCHARGE CHANNELS OF VARIOUS RADII, IMPINGING ON POLYTHENE SAMPLES IMMERSED IN B 30 TRANSFORMER OIL

Specimen thickness, t	Electric strength between British Standard electrodes, E_{av}	Radius of discharge channels, R	$E_{max} = 2E_{av}tp_1/\log q$	
mils	kV (r.m.s.) per mil	mils	kV (r.m.s.) per mil	MV (peak) per cm
3·5	3·15	1	9	5·0
		3	5·3	2·95
		5	4·6	2·6
		10	3·8	2·1
		20	3·5	1·95
62·5	0·48–0·60	1	10·9–13·5	6·1–7·5
		3	4·6– 5·8	2·6–3·25
		5	3·2– 4·0	1·8–2·2
		10	2·0– 2·5	1·1–1·4
		20	1·2– 1·6	0·7–0·9
250	0·20	1	14·5	8·3
		3	5·7	3·2
		5	3·8	2·1
		10	2·2	1·2
		20	1·3	0·73

calculated from step-by-step measurements* at 20° C on polythene discs between British Standard electrodes in oil. Values for R of between 1 and 20 mils are assumed. Table 3 shows that with the thicker specimens even the high-temperature intrinsic strength of polythene (1·5 MV/cm) would not be exceeded if the channels were of 10- or 20-mil radius; while the low-temperature intrinsic strength (6·5 MV/cm) would be exceeded with channels of 1-mil radius. The most consistent values of E_{max} (2–3 MV/cm) are given for R = 3–5 mils, so that it should be possible to predict the variation of step-by-step electric strength of polythene, with specimen thickness, from eqn. (4), assuming that E_{max} must exceed some 2–3 MV/cm at the end of channels of 3–5-mil radius.

(5.1) Variation of Industrial Electric Strength with Specimen Thickness

The curves in Fig. 10 have been calculated for several possible values of E_{max} and R. The best agreement with the experimental values for polythene tested between British Standard electrodes under oil is given by curve B for R = 3 mils and E_{max} = 5·25 kV (r.m.s.) per mil. The assumption that R is constant for all values of t may be an over-simplification, and in practice the value of R probably increases slightly with increasing specimen thickness. Knowledge of the exact values of R and E_{max} is not critical, however, as is shown by the proximity of the curves in Fig. 10.

Curve B is reproduced again in Fig. 11 together with measured industrial electric-strength values for polythene, mica,[14] Pyrex glass,[19] and ebonite,[15,16] which follow approximately the same

* Measured with the specimen in air or under oil, between standard electrodes, such as specified by the B.S.I., A.S.T.M., etc.

† A similar relation was postulated by A. E. W. Austen, in an unpublished note in 1946.

* The step-by-step breakdown voltage was determined by applying two-thirds of the short-time breakdown voltage for 1 min and then raising the voltage by steps of 1 or 2 kV/min until breakdown occurred.

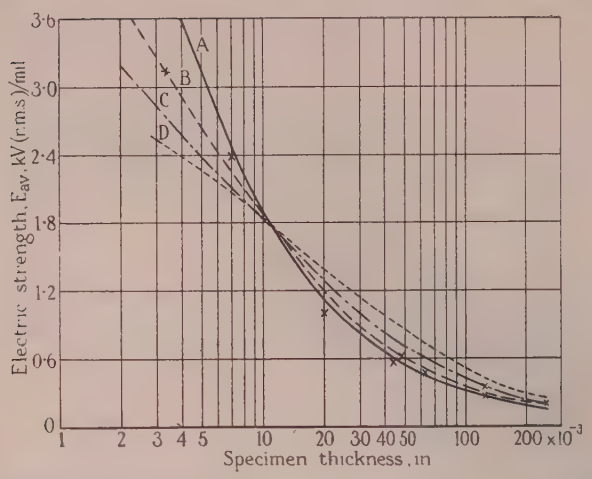


Fig. 10.—Variation of average breakdown strength with specimen thickness.

	R mils	E_{max} kV(r.m.s.)/mil
———	1	11
— — —	3	5·25
—·—·—	5	3·4
- - - - -	10	3·0
× × × ×	Experimental values for polythene (B.S.I. electrodes in oil).	

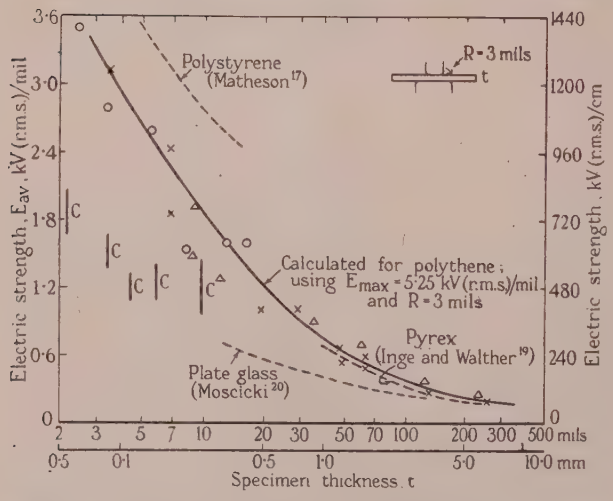


Fig. 11.—Variation with thickness of 1 min industrial electric strength of dielectric materials measured in clean oil at 20°C.

———	Measured values.
× × × ×	Polythene.
△ △ △ △	Ebonite.[15,16]
○ ○ ○ ○	Ruby mica.[14]
|C	Cellulose acetate at 90° C.[18]

curve. Curves for plate glass[13] and polystyrene[17] are of similar shape but displaced respectively below and above the calculated curve. Electric-strength values for cellulose acetate[18] are too scattered to show any correlation with thickness.

The similarity between these measured and calculated values is sufficiently striking to justify further investigation of the dependence of industrial electric-strength on sample thickness. Tests on materials free from defects and having low dielectric loss, to avoid breakdown by thermal instability, must be made under carefully controlled conditions, because the apparent electric strength of a material is determined not only by its own physical properties, but also by those of the medium, the electrode configuration, and by the waveform and total number of voltage cycles applied.[2,3] Table 4 shows a limited number of breakdown values obtained under various conditions in the course of other investigations, and allows some comparison to be made of the effect of the immersion medium and electrode configuration.

(5.2) Variation of Electric Strength with Discharge Inception Voltage

The variation of the electric strength with the inception voltage of discharges in the medium can be explained by considering both the stress concentration at the ends of the discharge channels and the energy of the individual discharges. If the inception voltage, V_i, is low, the stress at the ends of the channels is also low, and breakdown in short times can occur only at voltages greatly exceeding V_i; e.g. Table 4 shows that, in air, failure occurs at between 10 and 20 times the inception voltage. If V_i is greater, however, there is greater stress at the end of the discharge channels, and breakdown channels propagate through the material when the applied voltage is between 1·5 and $3V_i$, as shown in Table 4, using clean transformer oil as the medium. An increase in the value of V_i, by changing the immersion medium, may thus decrease the breakdown voltage of thick homogeneous materials, provided of course that V_i is not increased to a value exceeding the breakdown voltage in the former medium.

The energy of the discharges is important, because this will affect the local temperature rise at the point of impact of the discharges, and thus the value of intrinsic strength which must be attained. In Section 9.2 it is shown that the discharge energy per unit area is given by

$$W/A = E_M^2 l[\epsilon_D l + \epsilon_M t]/8\pi t = \epsilon_D^2 l V_i^2/8\pi t[\epsilon_D l + \epsilon_M t]$$

where E_M is the electric strength of a length l of the medium;

ϵ_M is the permittivity of the medium;
ϵ_D is the permittivity of the specimen of thickness t.

The value of W/A cannot be readily evaluated, because of uncertainty in the value of l; moreover, when the applied voltage is increased above V_i, discharges of greater length and greater energy may occur than at V_i. The value of V_i is nevertheless significant, as shown by Table 4, where discharges in oil, with V_i between 7 and 14 kV (r.m.s.), cause breakdown at much lower voltages than discharges in air, where V_i varies between 1 and 4 kV (r.m.s.) respectively for samples of the same thickness. Intermediate values, both of inception voltage and of electric strength are recorded, with contaminated oil, as shown in Table 4.

(5.3) Variation of Electric Strength with Time of Voltage Application

The relation between the life of insulation and the ratio of the applied (alternating) voltage, V, to the discharge inception voltage, V_i, is readily explained if cumulative erosion precedes failure,[1] but only tentative explanations are possible if breakdown channels propagate without preliminary erosion.[3]

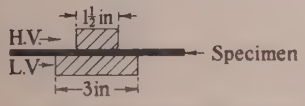
When the local stress initially exceeds the intrinsic electric strength (the high-temperature value, if the discharges cause local heating) the channels probably propagate through a few microns only, because the stress decreases rapidly with distance from the tip, as shown in Fig. 8 for a needle-point electrode, and the heating effect is very localized. Thus a succession of discharges must recur at the same site if the channels are to extend and cause failure. The time to breakdown will thus depend on the local stress concentration in relation to the intrinsic strength, the discharge energy, the discharge repetition frequency and the specimen thickness.

Table 4

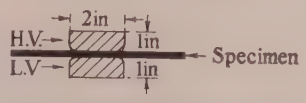
FACTORS AFFECTING THE A.C. STEP-BY-STEP ELECTRIC STRENGTH OF INSULATION

Electrodes	Sample thickness	Test medium					
		Air		Clean B30 transformer oil		Contaminated oil	
		V_i	Breakdown voltage	V_i	Breakdown voltage	V_i	Breakdown voltage
	mm	kV (r.m.s.)	kV (r.m.s.)	kV (r.m.s.)	kV (r.m.s.)	kV (r.m.s.)	kV (r.m.s.)
Perspex Specimens:							
A.S.T.M.	1	2·1	52·2 after 15sec	9	25 after 10sec	—	—
B.S.I.	1	—	—	10	22·5 after 55sec	—	—
Polystyrene Specimens:							
A.S.T.M.	1	3	60 after 20sec	13·5	40 after 10sec	6	52 when voltage increasing from 50
Polythene Specimens:							
B.S.I.	6·25	—	—	18	50 after 10sec	—	—
B.S.I.	3·12	—	—	—	35 after 10sec	—	42 after 35sec
A.S.T.M.	1·55	2·6	50 after 25sec	14	37·5 after 60sec	8	47·5 when voltage increasing from 45
B.S.I.	1·55	4·1	(*a*) 42; (*b*) 50 after 40sec	13·5	30 after 10sec	—	—
B.S.I.	1·2	3·1	30 after 300sec	16	30 after 20sec	—	—
B.S.I.	1·1	—	—	15	25 after 20sec	—	—
B.S.I.	0·5	2·3	20 after 200sec	12·5	20 after 20sec	—	—
B.S.I.	0·18	2·0	13 after 15sec	8	11 after 50sec	—	12·5 after 44sec
B.S.I.	0·09	1·6	11 after 10sec	7	11 after 15sec	—	—
Cellulose-Acetate Specimens:							
A.S.T.M.	0·25	0·9	14 after 30sec	8·25	9 after 30sec	7·25	11 when voltage increasing
A.S.T.M.	0·12	0·4	13·5 after 10sec	6·25	7 almost instantaneosuly	—	12·5 after 5sec

V_i = Discharge inception voltage.
Breakdown voltage, determined after previous "minute" tests at lower voltage.



B.S.I. electrode system (asymmetrical).



A.S.T.M. electrode system (symmetrical).

In thin specimens the average breakdown stress approaches the intrinsic electric strength, as shown in Fig. 11, so that the field at the end of discharge channels is much less divergent than it would be with thicker specimens. Discharge channels can thus propagate more rapidly in thin than in thick specimens, because there is less necessity for re-concentration of electric stress after initial penetration. This may explain why, in industrial electric-strength tests, partial-breakdown channels are often detected in thick specimens, where the average breakdown strength is less than one-tenth of the intrinsic value, but in thin specimens there is either no observable damage or complete breakdown. Partial-breakdown channels are also less often found when samples are tested in air than when under oil. This also may be due to the higher average electric stress (see Table 4) required for breakdown in air; i.e. the field at the end of the channels is less divergent when the specimen is tested in air than under oil. Breakdown in air is thus more rapid than under oil, so that there is less chance of incomplete-breakdown channels being observed, except with chopped impulse tests.

Correlation between the energy of discharges and the short-time electric strength of materials can be inferred from data on the breakdown of oil-impregnated and Askerel-impregnated pressboard, shown in Fig. 12.[23]

The long-time breakdown voltage tends asymptotically to V_i, so that it may be inferred from Fig. 12 that V_i with Askerel impregnant is some 50% greater than with oil impregnant. Thus with Askerel impregnant the energy of individual discharges is more than twice that with oil impregnant; the short-time breakdown voltage with Askerel impregnant will therefore be lower than with oil impregnant, as shown in Section 5.2.

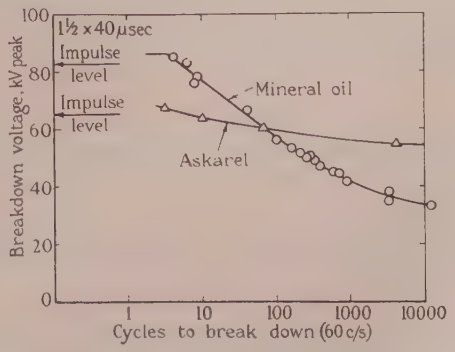


Fig. 12.—Time to break down impregnated pressboard (1/32in) (Dakin and Works).[23]

This effect may be of considerable practical importance because, although the alternating working voltage is increased if the discharge inception voltage is increased, the impulse electric strength may be decreased.

(5.4) The Effect of Polarity on the Impulse Electric Strength of Insulation

Fig. 6 shows that with point electrodes of radius between 1·5 and 7·5 microns and electrode spacings of between 250 and 800 microns, the electric strength of polythene with a negative point is some 25% greater than with a positive point at 20°C. This polarity effect is greater than any previously recorded with solid dielectrics; e.g. tests between point and plane electrodes show that the impulse electric strength of bitumen[20] is between 2 and 10% greater with negative than positive points. In alkali-halide crystals[21] breakdown channels propagate more rapidly at a given stress with positive than with negative points, but the ratio of the impulse breakdown stresses is not stated.

With impregnated-paper cables,[22] where breakdown is probably caused by internal discharges, the polarity effect is reversed, e.g. the impulse electric strength is about 10% greater with the inner conductor positive than with the opposite polarity. The explanation is probably as follows: In cavities between a dielectric and a metal surface,[1] a few large (high-energy) discharges occur when the metal is negative, and many smaller discharges when the dielectric surface is negative. Failure can thus occur at a lower stress when the cable inner conductor is negative than with reverse polarity, because breakdown channels propagate at a lower stress in the presence of high-energy discharges than with lower-energy discharges, as discussed in Section 5.2.

Space charge in the dielectric is unlikely to affect the electric strength in the presence of discharges, because the radius of the discharge channel, usually about 75 microns (3 mils), exceeds the radius of the space charge, which (in polythene) is some 25 microns with positive points and about 45 microns with negative points, as shown in Section 4.3.

(6) CONCLUSIONS

The propagation of breakdown channels from steel-point electrodes embedded in polythene to adjacent plane electrodes shows that breakdown can occur when the effective maximum stress attains the intrinsic electric strength of the material.

The variation of the divergent-field strength with the polarity and radius of curvature of the point indicate that stress-enhanced conductivity and space-charge accumulation reduce the stress near sharp points to values much lower than those calculated geometrically.

The variations of the industrial electric strength of insulation with specimen thickness, and with the electrical characteristics of the specimen and of the medium, are explained by the conditions necessary for breakdown in divergent fields.

(7) ACKNOWLEDGMENTS

The author wishes to acknowledge helpful discussions with his colleagues, and to thank Mr. G. A. Prinz for practical assistance.

Acknowledgment is due to the Director of the British Electrical and Allied Industries Research Association for permission to publish the paper.

(8) LIST OF REFERENCES

(1) MASON, J. H.: "The Deterioration and Breakdown of Dielectrics resulting from Internal Discharges," *Proceedings I.E.E.*, Paper No. 1053, January, 1951 (**98**, Part I, p. 44.

(2) MASON, J. H.: "Breakdown of Insulation by Discharges," *ibid.*, Paper No. 1471 M, March, 1953 (**100**, Part IIA, p. 149).

(3) OGILVIE, K. W.: E.R.A. Report (in course of preparation).

(4) WRIGHT, D. A. "A Survey of Present Knowledge of Thermionic Emitters," *Proceedings I.E.E.*, Paper No. 1404 R, November, 1952 (**100**, Part III, p. 125).

(5) HUGHES, A. L., and DU BRIDGE, L. A.: "Photo-Electric Phenomena," 1932 (McGraw-Hill Book Co., New York).

(6) SALVAGE, B.: "The Dielectric Breakdown of Some Simple Organic Liquids," *Proceedings I.E.E.*, Monograph No. 2, May, 1951 (**98**, Part IV, p. 1).

(7) GOODWIN, D. W., and MCFADYEN, K. A.: "Electrical Conduction and Breakdown in Liquid Dielectrics," *Proceedings of the Physical Society*, 1953, **66**, p. 85.

(8) CALDERWOOD, J. H., COOPER, R., and WALLACE, A. A.: "An Experimental Investigation of Factors influencing the Electric Strength of Potassium Chloride Crystals," *Proceedings I.E.E.*, Paper No. 1457 M, March, 1953 (**100**, Part IIA, p. 105).

(8A) COOPER, R., and GROSSART, D. T.: "The Influence of Cathode Material on the Electric Strength of Potassium Bromide Crystals," *Proceedings of the Physical Society*, 1953, **66** B, p. 716.

(9) HAWLEY, E. M.: "The Variation with Temperature of the Electric Strength of Polyisobutylene subjected to Discharges," E.R.A. Report Ref. L/T290, 1953.

(10) WHITEHEAD, S.: "Dielectric Breakdown of Solids" (Clarendon Press, 1951), Chapter 1.

(11) WHITEHEAD, S.: "Breakdown of Solid Dielectrics" (Benn, 1932), Chapter 3.

(12) LIAO, T. W., and ANDERSON, J. G.: "Propagation Mechanism of Impulse Corona and Breakdown in Oil," *Communication and Electronics*, 1953, No. 9, p. 641.

(13) MOSCICKI: *Electrotechnische Zeitschrift*, 1904, **25**, p. 527.

(14) LEWIS, A. B., HALL, E. L., and CALDWELL, F. R.: "Some Electrical Properties of Micas," *Journal of Research of the National Bureau of Standards*, 1931, **7**, p. 403.

(15) FARMER, F. M.: "Dielectric Strength of Thin Insulating Materials," *Transactions of the American I.E.E.*, 1913, **32**, p. 2097.

(16) PEEK, F. W.: "Dielectric Phenomena in High Voltage Engineering" (New York, 1916).

(17) MATHESON, L. A., and GOGGIN, W. C.: "Electrical Application of Polystyrene," *Industrial and Engineering Chemistry*, 1939, **31**, p. 334.

(18) "Cellulose Derivative Films for Insulating Purposes," E.R.A. Report Ref. A/T118, 1950.

(19) INGE, L., and WALTHER, A.: *Archiv für Electrotechnik*, 1930, **24**, p. 88.

(20) STANDRING, W. G., and CHAPPEL, D. W. G.: "The Breakdown of Bitumen under Impulse Voltages," E.R.A. Report Ref. L/T113, 1943.

(21) INGE, L., and WALTHER, A.: "Elektrische Entladungen Kristallen," *Zeitschrift für Physik*, 1930, **64**, p. 830.

(22) HOWARD, P. R.: "Impulse Puncture Characteristics of Mass-Impregnated Paper-Insulated Cables," *Proceedings I.E.E.*, Paper No. 1515 S, June, 1953 (**100**, Part II, p. 315).

(23) DAKIN, T. W., and WORKS, C. N.: "Impulse Dielectric Strength Characteristics of Liquid Impregnated Pressboard," *Transactions of the American I.E.E.*, **71**, Part I, p. 321.

(24) FLYNN, P.: "Field Dependent Conductivity in Non-Uniform Field Distributions and its Relation to Electrical Breakdown" (see page 264).

(25) HUNTER, E., and OAKES, W. G.: "The Effect of Temperature on the Density of Polythene," *Transactions of the Faraday Society*, 1945, **41**, p. 49.

(9) APPENDICES

(9.1) Method of Correction for Thermal Expansion of Polythene

When the polythene discs are heated in the plate electrode, the centre melts, but the periphery, which projects beyond the electrode, is only slightly heated. The area of the disc is thus fixed, so that the expansion of the polythene occurs only as an increase in specimen thickness, t'. At the melting point $t'_{120°C} \simeq 1{\cdot}15t'_{20°C}$. If the needle is inserted to a depth c, the distance between the point and plane, at 120°C is $t_{120} = (1{\cdot}15t'_{20} - c)$. The polythene adheres to the needle, which is not clamped, so that, when the polythene contracts on cooling, the distance between point and plane, at a temperature T is

$$t_T = (1{\cdot}15t'_{20} - c)[1 - (120 - T)\gamma]$$

where γ is the coefficient of cubical expansion of polythene. γ varies with temperature as follows:[25]

20 to 50°C $\gamma = 0{\cdot}0007$ per °C
50 to 80°C $\gamma = 0{\cdot}001$ per °C
80 to 120°C $\gamma = 0{\cdot}0028$ per °C

At 20°C, $t_{20} = (t'_{20} - c \times 100/115)$, and this value has been checked experimentally.

All the values of t and the derived stress values have been corrected for the effects of thermal expansion.

(9.2) Energy Liberated by a Discharge

A sample between industrial electrodes can be represented by the circuit in Fig. 13.

Provided that C_A is large compared with C_B and C_C, only the element of edge capacitance associated with each discharge need be considered in the subsequent calculations.

The energy W liberated when the voltage across C_C changes by δV_C, is given by

$$2W = [C_A C_B/(C_A + C_B) + C_C][\delta V_C^2 - 2\delta V_C V_C] \quad . \quad (9)$$

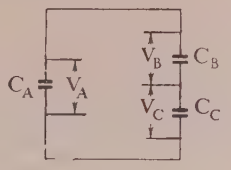


Fig. 13.—Circuit equivalent to sample between test electrodes.
C_A = Capacitance of specimen between electrodes.
C_B = Edge capacitance of specimen.
C_G = Edge capacitance of the air in series with the specimen.
V_A, V_B and V_G = Voltages across the capacitances.

If C_C is almost completely discharged, $\delta V_C \simeq V_C$, and if $C_A > C_B \simeq C_C$,

$$2W = (C_B + C_C)V_C^2 = C_B^2 V_A^2/(C_B + C_C) \quad . \quad . \quad (10)$$

At the instant of a discharge

$$V_A = V_i = E_M(\epsilon_D l + \epsilon_M t)/\epsilon_D \quad . \quad . \quad . \quad (11)$$

where V_i = Discharge inception voltage.

E_M = Electric strength of a length l of the medium.

ϵ_M = Permittivity of the medium.

ϵ_D = Permittivity of the sample of thickness t.

And where the capacitances C_B and C_C are given by eqns. (12) and (13).

$$C_B = \epsilon_D A/4\pi t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

$$C_C = \epsilon_M A/4\pi l \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

(assuming that an area A is affected by the discharge).

If these expressions for C_B, C_C and V_A are substituted in eqn. (10) the discharge energy dissipated per unit area is found to be

$$W/A = \epsilon_D^2 l V_i^2/8\pi t(\epsilon_D l + \epsilon_M t) = E_M^2 l[\epsilon_D l + \epsilon_M t]/8\pi t$$

621.3.015.5: 621.3.013

Monograph No. 128 M
Apr. 1955

FIELD-DEPENDENT CONDUCTIVITY IN NON-UNIFORM FIELDS AND ITS RELATION TO ELECTRICAL BREAKDOWN

By PATRICK T. G. FLYNN, B.Sc.(Eng.), Ph.D., Graduate.

(*The paper was first received* 3*rd November*, 1954, *and in revised form* 11*th January*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in April*, 1955.)

SUMMARY

A method is given for calculating the resulting field distribution when the conductivity is a function of the field strength. This is applied to a point-plane electrode system in an amorphous solid, with the assumption that the conductivity at low field-strengths is electronic and that the conductivity at high field-strengths is described by a relation due to Fröhlich. The results show that for polymethylmethacrylate with average fields of the order of 10^6 volt/cm, the reduction in maximum field-strength from that calculated for constant conductivity can be a factor of five or more, the magnitude of the reduction depending on the degree of non-uniformity of the field. Energy-level data required for the calculation are given for glass, polyethylene, polymethylmethacrylate and several types of varnish, and incomplete data are given for chlorinated polyethylene, polystyrene and polyisobutylene. The results are applied to breakdown in a non-uniform field, and a method is suggested for determining experimentally whether breakdown occurs when the maximum stress reaches the intrinsic electric strength, or whether it is necessary for the intrinsic electric strength to be exceeded over a given distance.

LIST OF PRINCIPAL SYMBOLS

a = Parameter used in defining geometry of electrodes.
E = Field strength, volts/cm.
E_i = Intrinsic breakdown field, volts/cm.
E_t = Transformed field-strength, volts/cm.
J = Current density, amp/cm^2.
k = Boltzmann's constant.
R = Radius of curvature of the end of point electrode, cm.
σ = Conductivity, mho/cm.
σ_0 = Conductivity in very weak fields, mho/cm.
T = Temperature, °K.
V = Potential, volts.
V_0 = Potential difference between electrodes, volts.
V_t = Transformed potential, volts.
$(V_t)_0$ = Transformed potential difference between electrodes, volts.
W = Energy difference between the conduction band and electrons in deep traps, eV.
ΔW = Energy range of electrons in shallow traps immediately below the conduction band, eV.
x = Co-ordinate.
x_0 = Distance from plane electrode to end of point electrode, cm.
x_1 = Distance over which the intrinsic electric strength is exceeded, cm.
y = Co-ordinate.

(1) INTRODUCTION

The calculation of the field distribution for an electrode configuration producing a non-uniform field is normally made with the assumption that the conductivity of the material between the electrodes is constant. At high field-strengths, however, the conductivity may become field dependent, and the resulting distribution will not then be the same as at low field-strengths.

It appears to be generally accepted that in insulating solids the conductivity increases when very high field-strengths are applied. In the case of a point-to-plane electrode system with solid dielectric, a high field-strength at the point will be reduced as a result of enhanced conductivity, and it therefore becomes of interest to know what the modified field distribution will be; this is of particular importance in work on the electrical breakdown of dielectrics in non-uniform fields.

It is the purpose of the present paper to consider the effect of field-dependent conductivity in non-uniform fields, and to derive a method for the calculation of the field distribution. The method will be applied to an amorphous dielectric with a point-to-plane electrode system.

(2) THE FIELD EQUATIONS

Consider the general case of current flow. For the steady-state condition

$$\text{div}\, J = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where J is the current-density vector. If the field-strength vector is E and the conductivity is σ, then

$$J = \sigma E \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Combining eqns. (1) and (2) gives

$$\text{div}\,(\sigma E) = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Now if σ is not a function of position and is not a function of E, and if E is related to the potential V by the relation

$$E = -\,\text{grad}\, V \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

eqn. (3) becomes

$$\text{div grad}\, V = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

This is Laplace's equation, and it can be solved for V for cases of interest.

Consider now the case when σ is a function of the field E; the solution of eqns. (3) and (4) then becomes more difficult. However, a solution can be obtained by the following transformation. Denote by E_t a function of the field E defined by

$$E_t = (\sigma/\sigma_0)E \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where σ_0 is the conductivity at some standard field. Eqn. (3) then becomes

$$\text{div}\,(\sigma_0 E_t) = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

A second potential function will be denoted by V_t where

$$E_t = -\,\text{grad}\, V_t \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Since σ_0 is a constant, eqn. (7) then becomes

$$\text{div grad}\, V_t = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Thus any field distribution E and associated potential distribution V for which the conductivity is a function of the field can be replaced by a transformed field-distribution E_t and its associated potential distribution V_t for which the conductivity is constant.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

This paper is based on Report L/T319 of the British Electrical and Allied Industries Research Association.

Dr. Flynn is now at the Atomic Weapons Research Association.

Fortunately, the transformed potential satisfies Laplace's equation, as in eqn. (9), and the original field distribution can therefore be found for any geometry for which solutions of Laplace's equation are known.

The procedure is as follows: For the given geometry with applied potential difference $(V_t)_0$, the field distribution E_t is calculated. By means of eqn. (6) this is transformed to a field distribution E, and by integration the value of the corresponding applied potential difference V_0 is found. This is repeated for several values of $(V_t)_0$ and a graph of corresponding values of $(V_t)_0$ and V_0 is plotted. The field E corresponding to a given value of V_0 can then be found by calculating the E_t distribution for the correct value of $(V_t)_0$ taken from the graph, and transforming from E_t to E by use of eqn. (6).

It may be noted that if σ is field dependent, then for the non-uniform field σ will also be a function of position. Hence from eqn. (3)

$$\operatorname{div} E = -(E/\sigma) \operatorname{grad} \sigma \quad . \quad . \quad . \quad . \quad (10)$$

i.e. space charge will be present unless σ is constant.

(3) ELECTRICAL CONDUCTIVITY AT HIGH FIELD-STRENGTHS IN AMORPHOUS SOLIDS

Theories for the increase of conductivity with field strength in dielectrics have been discussed by several authors and these have been reviewed by Whitehead.[15] The only theory immediately applicable to amorphous dielectrics is that due to Fröhlich.[3]

Fröhlich assumes that the distribution of the energy levels in an amorphous solid is similar to that of a crystal with many lattice imperfections, and the validity of this assumption has been discussed by Simpson.[10] On the basis of this model the dependence of the conductivity on field strength is given by

$$\log(\sigma/\sigma_0) = \left\{ \begin{array}{ll} WE^2/2{\cdot}72\Delta W E_i^2 & \ldots \text{if } E^2 \ll E_i^2 \\ W/\Delta W & \ldots \text{if } E^2 \simeq E_i^2 \end{array} \right\} . \quad (11)$$

W can be found from the temperature dependence of the conductivity in very weak fields assuming that

$$\sigma_0 = (\text{constant}) \exp(-W/kT) \quad . \quad . \quad . \quad . \quad (12)$$

ΔW can be found from the temperature dependence of the breakdown field, which according to Fröhlich[3] is given by

$$\log E_i = \text{constant} + \Delta W/2kT \quad . \quad . \quad . \quad (13)$$

in the high-temperature region of breakdown.

The experimental evidence for the increase of conductivity at high field-strengths is limited because of the difficulty of making measurements. The only measurements with maintained voltage are those of Thomas and Griffith[11] on varnish films; these show an increased conductivity, but, as noted by Fröhlich, $\log(\sigma/\sigma_0)$ is not exactly proportional to E^2. Turner and Lewis[13] measured current/voltage characteristics of soda-lime glass using a linearly rising voltage pulse of 150 microsec duration and found that the conductivity increased with applied field according to Fröhlich's equation, although the various anomalous currents arising in measurements of this kind may invalidate direct use of the results. Similar measurements on soda-lime glass using square voltage pulses of 30 microsec and 110 microsec have been made by Tomura and Kikuchi;[12] the results, which exclude charging currents, verify the dependence of $\log \sigma$ on E^2.

(3.1) Application to Polymethylmethacrylate

To apply the above analysis to an example of an amorphous dielectric, it is necessary to know the values of ΔW and W in eqn. (11). From the point of view of the breakdown measurements described in an accompanying paper,[5] it would be desirable to make the calculation for polyethylene, but the available data for this material, as shown in Table 1, show a range of values for W, and also give an overlap between W and ΔW such that the model leading to eqn. (11) may no longer apply. For this reason it has been thought better to illustrate the principles involved by calculations for polymethylmethacrylate, for which the data in Table 1 appear more suitable.

Oakes[7] has measured the intrinsic electric strength of polymethylmethacrylate over a wide range of temperature, and from the slope of the graph of $\log E_i$ against T^{-1} [see eqn. (13)] values of ΔW between 0·16eV and 0·195eV can be obtained,* according to the position assumed for the transition to the high-temperature region of electrical breakdown.

Measurements of the temperature dependence of the conductivity in weak fields for polymethylmethacrylate have been made by Fowler and Farmer[2] over a rather limited temperature range. The use of eqn. (12) gives a value of 0·5eV for W.

In order to simplify the calculation it will be assumed that the part of eqn. (11) applying for $E^2 \ll E_i^2$ is valid also in the region where $E^2 \simeq E_i^2$; to balance the effect of this ΔW will be chosen at the lower end of its range, i.e. $\Delta W = 0{\cdot}16$eV.

From Oakes's results, the intrinsic electric strength E_i at a temperature of 16°C is $11{\cdot}4 \times 10^6$volt/cm.

Inserting the numerical values in eqn. (11) gives

$$\log(\sigma/\sigma_0) = 8{\cdot}75 \times 10^{-15} E^2 \quad . \quad . \quad . \quad . \quad (14)$$

and hence the required transformation is obtained from eqn. (6) as

$$E_t = E \exp(8{\cdot}75 \times 10^{-15} E^2) \quad . \quad . \quad . \quad (15a)$$

Similarly, at temperatures of 50°C and 100°C, the transformations are

$$E_t = E \exp(18{\cdot}7 \times 10^{-15} E^2) \quad . \quad . \quad . \quad (15b)$$

and

$$E_t = E \exp(48{\cdot}3 \times 10^{-15} E^2) \quad . \quad . \quad . \quad (15c)$$

respectively. Eqn. (15) is plotted in Fig. 1.

(3.2) Energy Levels in Some Insulators

A list of the values obtained for the energy levels in some insulating materials is given in Table 1.

It should be noted that the values of W may be uncertain. Eqn. (12), which has been used to calculate values of W, is based on the assumption that the conduction is electronic, and it will therefore be in error if the observed conductivity is ionic in nature.†

Absorption current will also affect the result if measurements are made with short times.

The values of ΔW obtained by use of eqn. (13) depend on the temperature selected as representing the transition from the low-temperature region to the high-temperature region of breakdown. This temperature is reasonably well defined for the materials listed. For soda-lime glass the value of 0·65eV for ΔW found by Tomura and Kikuchi is obtained, not from the temperature variation of the breakdown field, but from the temperature variation of $\log(\sigma/E^2)$ using eqns. (11) and (13).

(4) THE FIELD DISTRIBUTION FOR A POINT-TO-PLANE GEOMETRY

For a point-to-plane geometry in which the "point" is a hyperboloid of revolution, the equipotential surfaces are also

* Oakes quotes a value for ΔW of 0·06eV, but this appears to be in error. Oakes plots values of $\log_{10} E_i$, and the factor 2·3 to convert to $\log E_i$ has apparently been omitted in the calculation.

† Mayburg and Lawrence[6] suggest that the conductivity in polyethylene may be ionic, the current carrier being the proton. However, the theoretical argument in support of this does not appear to be based on valid assumptions, and it seems more probable that the conductivity is electronic.

Table 1

ENERGY LEVELS

Material	Energy level		Source
		eV	
Black baking bituminous varnish	W	2·7	Thomas and Griffith[11] (1942)
	ΔW	0·06	
Clear baking oil insulating varnish	W	2·0	
	ΔW	0·13	
Bakelite "formite" varnish	W	3·1	
	ΔW	0·035	
Soda-lime glass	W	0·67	Tomura and Kikuchi[12] (1952)
		0·5	Simpson[10] (1950)
	ΔW	0·57 0·65	Tomura and Kikuchi[12] (1952)
		0·29	Von Hippel and Maurer[14] (1941)
Polyethylene	W	0·5	Ramsey[9] (1953)
		0·27	Mayburg and Lawrence[6] (1952)
	ΔW	0·46	Oakes[7] (1949)
Chlorinated polyethylene	W	No data available	
	ΔW	0·046	Oakes[7] (1949)
Polymethylmethacrylate	W	0·5	Fowler and Farmer[2] (1954)
	ΔW	0·16	Oakes[7] (1949)
Polystyrene	W	No data available	
	ΔW	0·46	Oakes[7] (1949)
Polyisobutylene	W	No data available	
	ΔW	0·41	Oakes[7] (1949)

hyperboloids of revolution. The field distribution for such a geometry has been derived by Eyring, Mackeown and Millikan.[1]

If the point surface is the hyperboloid generated by revolution of the hyperbola

$$\frac{x^2}{x_0^2} - \frac{y^2}{a^2 - x_0^2} = 1 \quad . \quad . \quad . \quad . \quad . \quad (16)$$

about the x-axis, the potential along the x-axis is

$$V_t = (V_t)_0 \frac{\log\,[(a + x)/(a - x)]}{\log[(a + x_0)/(a - x_0)]} \quad . \quad . \quad . \quad (17)$$

where V_0 is the potential at $x = x_0$ (i.e. the surface of the point), and the potential at $x = 0$ is taken as zero. Hence the field along the x-axis is

$$E_t = -\frac{2(V_t)_0}{a \log\,[(a + x_0)/(a - x_0)]}\left(\frac{a^2}{a^2 - x^2}\right) \quad . \quad (18)$$

Now for the hyperbola of eqn. (16), if R is the radius of curvature at the point $(x_0, 0)$ then

$$R = -x_0[(a^2/x_0^2) - 1] \quad . \quad . \quad . \quad . \quad (19)$$

Values of R/x_0 and a/x_0 calculated from eqn. (19) are shown in Fig. 2.

Thus measurement of R and x_0 for a given electrode system allows a to be found, and the field distribution can then be calculated from eqn. (18). It may be noted that eqn. (18) is exact; errors will arise if the point is not exactly a hyperboloid, but these will be small for cases where R is small compared to x_0.

(5) RESULTS FOR POINT-TO-PLANE ELECTRODE SYSTEM WITH POLYMETHYLMETHACRYLATE AS THE INSULATING SOLID

Fig. 3 shows a typical series of curves obtained by the method of Section 2. The curves are drawn for values of $(V_t)_0$ of 150kV, 100kV, 50kV and 10kV, the field being transformed by using Fig. 1, and the corresponding values of V_0 are found to be 75·1kV, 60·5kV, 38·7kV and 9·8kV as shown. For comparison, the broken curve shows the field distribution calculated for constant conductivity with $V_0 = 75{\cdot}1$kV. Similar curves are shown in Figs. 4 and 5.

From a series of such plots it is possible to obtain general curves for polymethylmethacrylate at 16°C relating values of

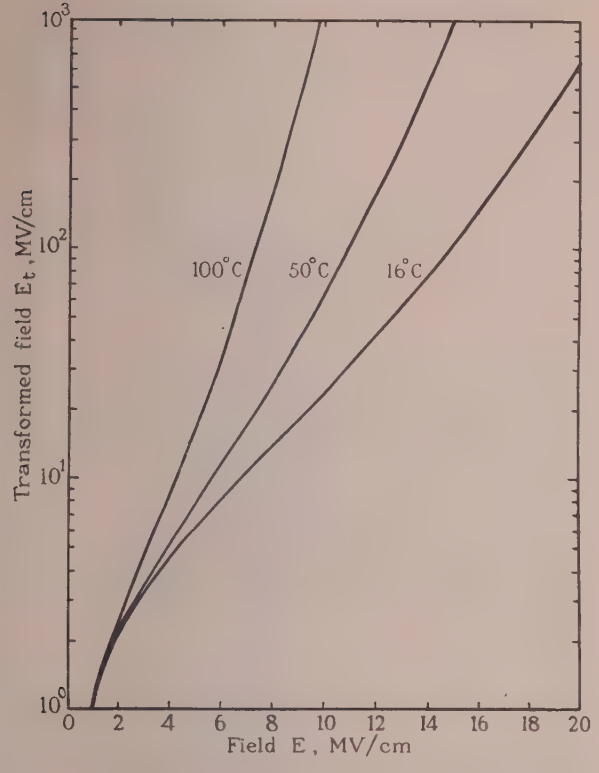


Fig. 1.—The function $E_t = (\sigma/\sigma_0)E$ for polymethylmethacrylate.

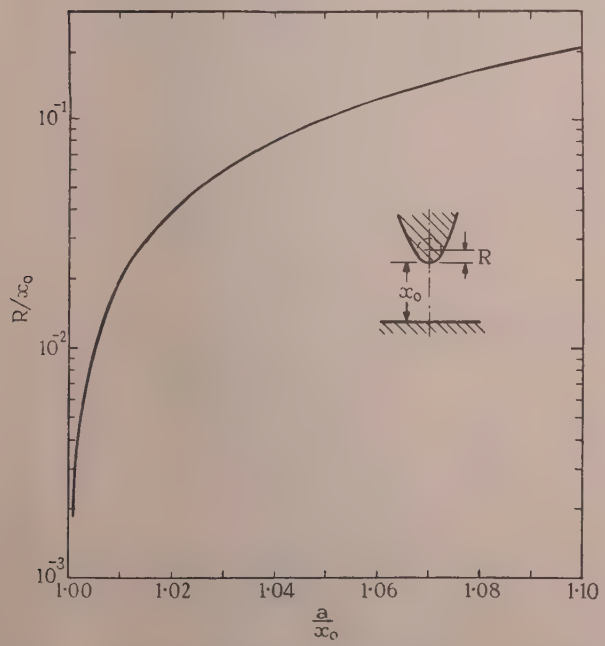


Fig. 2.—The parameter a for a hyperboloid of revolution in terms of R and x_0.

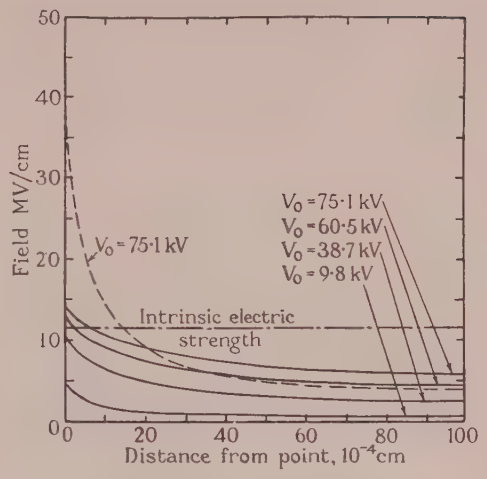


Fig. 3.—Field distribution for point with $R = 10^{-3}$cm and $x_0 = 10^{-2}$cm in polymethylmethacrylate at 16°C.
For comparison, broken curve shows field calculated for constant conductivity.

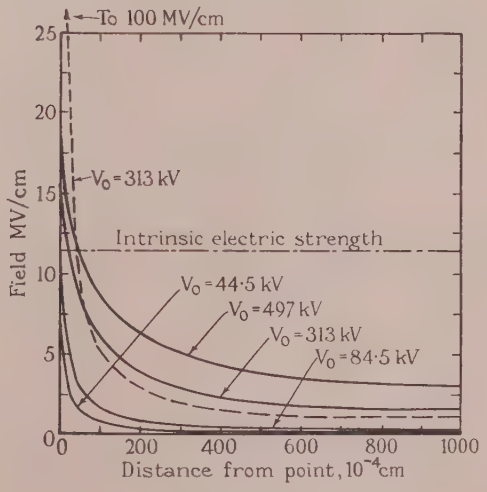


Fig. 4.—Field distribution for point with $R = 10^{-3}$cm and $x_0 = 10^{-1}$cm in polymethylmethacrylate at 16°C.
For comparison, broken curve shows field calculated for constant conductivity.

the reduced voltage $(V_t)_0$ and the actual voltage V_0 as shown in Fig. 6. It is easily seen by reference to eqns. (18) and (19) that all systems with the same values of $(V_t)_0/x_0$ and of R/x_0 are similar. Fig. 6 enables the correct value of $(V_t)_0$ to be chosen for any desired value of V_0.

Fig. 7 gives general curves for the maximum field-strength; these are compared with values calculated for constant conductivity.

(6) APPLICATION TO BREAKDOWN IN NON-UNIFORM FIELDS

An interesting application of the field distribution calculated above is the determination of the average field-strength required for breakdown with a point-to-plane geometry. Two possibilities exist: (*a*) that breakdown occurs when the maximum

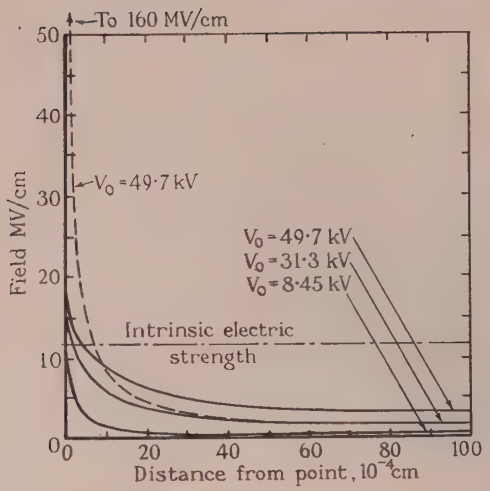


Fig. 5.—Field distribution for point with $R = 10^{-4}$cm and $x_0 = 10^{-2}$cm in polymethylmethacrylate at 16°C.
For comparison, broken curve shows field calculated for constant conductivity.

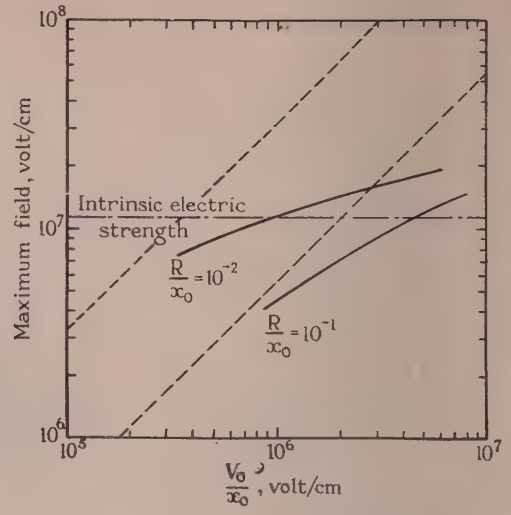


Fig. 7.—Maximum field for point-to-plane electrodes in polymethylmethacrylate at 16°C.
For comparison, broken curves show values for constant conductivity.

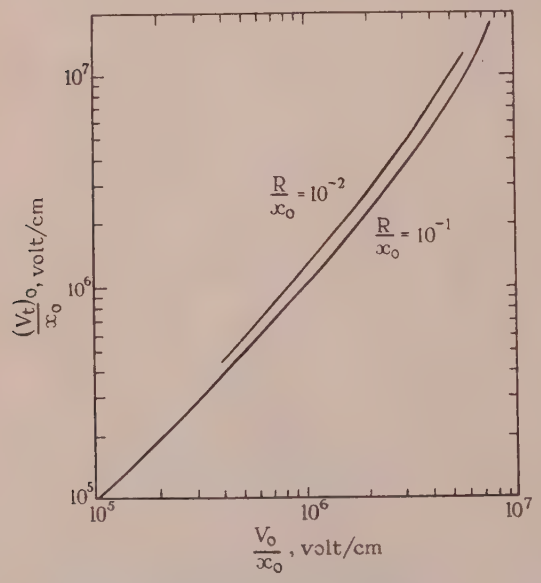


Fig. 6.—General curves for transformed voltage, $(V_t)_0$, in terms of applied voltage V_0 for point-to-plane electrodes in polymethylmethacrylate at 16°C.

stress reaches the intrinsic electric strength, and (*b*) that breakdown occurs when the intrinsic electric strength is exceeded over a given distance. For case (*a*) the average breakdown fields for polymethylmethacrylate at 16°C can be found from Fig. 7, and some values are given in Table 2. Fig. 8 shows the distance x_1 over which the intrinsic electric strength is exceeded for point-to-plane electrodes in polymethylmethacrylate at 16°C, and the values of the average field-strength for breakdown found from it—assuming that the required minimum distance is 10^{-4}cm—are given in Table 3.

These results show that there exists a possibility of distinguishing experimentally between cases (*a*) and (*b*) above in two ways. First, there is a difference in the magnitude of the average breakdown field-strength for the two cases. This difference may

Table 2

AVERAGE FIELD-STRENGTH REQUIRED FOR BREAKDOWN ON ASSUMPTION THAT MAXIMUM STRESS MUST EQUAL THE INTRINSIC ELECTRIC STRENGTH FOR POLYMETHYLMETHACRYLATE AT 16°C

R \ x_0	100 × 10^{-4}cm	1 000 × 10^{-4}cm
cm × 10^{-4}	MV/cm	MV/cm
1	1	
10	4	1
100		4

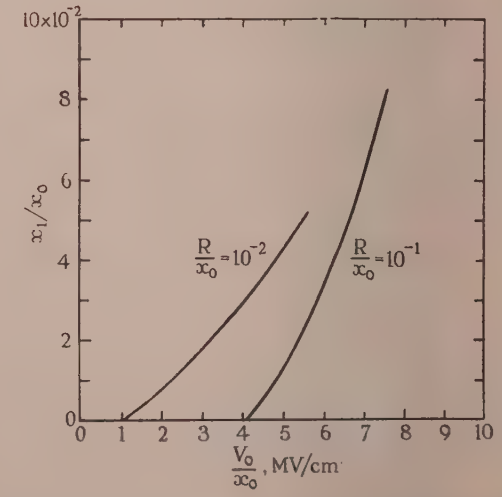


Fig. 8.—The distance x_1 over which the intrinsic electric strength is exceeded for point-to-plane electrodes in polymethylmethacrylate at 16°C.

Table 3

AVERAGE FIELD-STRENGTH REQUIRED FOR BREAKDOWN ON ASSUMPTION THAT THE INTRINSIC ELECTRIC STRENGTH MUST BE EXCEEDED OVER A DISTANCE OF 10^{-4} CM FOR POLYMETHYLMETHACRYLATE AT 16°C

R \ x_0	100×10^{-4}cm	1000×10^{-4}cm
cm × 10^{-4}	MV/cm	MV/cm
1	2·2	
10	4·75	1·2
100		4·2

be of the order of a factor of 2 for a suitable geometry (e.g. $R = 10^{-4}$cm and $x_0 = 100 \times 10^{-4}$cm). A second and more sensitive method of distinguishing between the two cases is to make a series of tests with a constant value of R/x_0. If breakdown occurs with the same average field, then the critical point is when the maximum stress reaches the intrinsic electric strength; if breakdown occurs with a decreasing average field as the physical size of the system increases, then the critical point is when the intrinsic electric strength is exceeded over a given distance. For best results a small value of R/x_0 is required. With polymethylmethacrylate, Table 3 shows that it should be possible to obtain a change of a factor of 2 in the average breakdown field-strength for a change of a factor of 10 in the size of the system when $R/x_0 = 10^{-2}$. If such a decrease is in fact obtained it should be possible to estimate the minimum distance required. Such tests would better be made with the point as anode, to eliminate, so far as possible, the effect of field or thermionic emission from the cathode into the region of high stress.

Field-emitted or thermionic electrons from a negative point may still further modify the field around the point and cause the maximum field to be lower than with a positive point (see O'Dwyer[8]). If this effect occurs the breakdown voltage for a negative point should be greater than that for a positive point, assuming that the same field conditions are required in each case; recent experiments by Mason[5] have shown that a polarity difference of this nature may exist for point-to-plane electrodes in polyethylene. Polarity effects may disappear at higher temperatures, since at the higher temperature the intrinsic electric strength decreases and the field in the solid necessary to initiate breakdown may be insufficient to cause appreciable emission from the cathode. Lewis[4] has suggested this as a reason for the non-appearance of cathode effects in some types of electrical breakdown in liquids.

(7) CONCLUSIONS

Field distributions for non-uniform geometries calculated in the normal manner may be considerably in error when high field-strengths are applied to insulating solids.

(8) ACKNOWLEDGMENT

Acknowledgment is made to the Director of the Electrical Research Association for permission to publish the paper, and to the author's colleagues for helpful discussion. That field-enhanced conductivity might be important in breakdown in non-uniform fields was suggested to the author by Dr. A. E. W. Austen.

(9) REFERENCES

(1) EYRING, C. F., MACKEOWN, S. S., and MILLIKAN, R. A.: "Field Currents from Points," *Physical Review*, 1928, **31**, p. 900.

(2) FOWLER, J. F., and FARMER, F. T.: "Conductivity induced in Insulating Materials by X-rays," *Nature*, 1954, **173**, p. 317.

(3) FRÖHLICH, H.: "Theory of Dielectric Breakdown in Amorphous Solids," *Proceedings of the Royal Society*, 1947, **A188**, p. 521.

(4) LEWIS, T. J.: "The Dependence of the Dielectric Strength of Pure Liquids on Cathode Material," *Proceedings of the Physical Society*, 1953, **B66**, p. 425.

(5) MASON, J. H.: "Breakdown of Solid Dielectrics in Divergent Fields," *Proceedings I.E.E.*, Monograph No. 127 M, April, 1955 (**102** C).

(6) MAYBURG, S., and LAWRENCE, W. L.: "The Conductivity Change in Polyethylene during Gamma-Irradiation," *Journal of Applied Physics*, 1952, **23**, p. 1006.

(7) OAKES, W. G.: "The Electric Strength of Some Synthetic Polymers," *Proceedings I.E.E.*, Paper No. 780, January, 1949 (**96**, Part I, p. 37).

(8) O'DWYER, J. J.: "The Influence of Cathode Material on the Measured Breakdown Strengths of Solid and Liquid Dielectrics," *Australian Journal of Physics*, 1954, **7**, p. 400.

(9) RAMSEY, N. W.: "The Effect of Temperature on the Conductivity induced in Insulators by X-rays," *Nature*, 1953, **172**, p. 214.

(10) SIMPSON, J. H.: "Time Delay in Conduction and Breakdown Processes in Amorphous Solids," *Proceedings of the Physical Society*, 1950, **A63**, p. 86.

(11) THOMAS, A. M., and GRIFFITH, M. V.: "Intrinsic Electric Strength and Conductivity of Varnish Films and their Variation with Temperature," *Journal I.E.E.*, 1942, **89**, Part I, p. 487.

(12) TOMURA, M., and KIKUCHI, T.: "Electronic Conductivities of Amorphous Solids at High Electric Fields," *Journal of the Physical Society of Japan*, 1952, **7**, p. 538.

(13) TURNER, C. H. M., and LEWIS, W. E.: "Electrical Conductivity in Insulating Amorphous Solids," *Nature*, 1947, **159**, p. 334.

(14) VON HIPPEL, A., and MAURER, R. J.: "Electrical Breakdown of Glasses and Crystals as a Function of Temperature," *Physical Review*, 1941, **59**, p. 820.

(15) WHITEHEAD, S.: "Dielectric Breakdown of Solids" (Clarendon Press, Oxford, 1951).

621.372.4/.5 : 621.3.012

Monograph No. 129 R
Apr. 1955

THE GRAPHS OF ACTIVE NETWORKS

By W. S. PERCIVAL, B.Sc., Associate Member.

(*The paper was received 27th November*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in April*, 1955.)

SUMMARY

The results obtained in previous papers are employed to develop the properties of the graphs of linear networks which may contain valves and transformers. A reduced graph is obtained which facilitates the setting up of the determinant of the network. This graph may be split into two separate graphs, termed the current and voltage graphs, such that there is a 1 : 1 correspondence between every algebraic operation on the *H*-matrix and every topological operation on the graphs. Thus every problem which can be solved with the one can be solved with the other. The derivative of a graph with respect to a network element corresponds to the derivative of an *H*-determinant with respect to a network element. With an extended definition of a tree on a network the set of trees on any network is shown to be equal to the nodal determinant of the network. Practical applications of the methods described are outlined.

LIST OF SYMBOLS

D = Ordinary determinant of network.
$f_{kj.im}$ = Fictitious element equal to transfer admittance $Y_{ij.km}$.
g = Mutual admittance of valve.
h = Mutual admittance of element $h_{ij.km}$.
h_i'' = Current-node admittance function for the node i.
h'_j = Voltage-node admittance function for the node j.
h_{ij} = Matrix or determinant element in row i and column j.
= Node-pair admittance.
$h_{ij.km}$ = Element of mutual admittance h with current nodes i and k and voltage nodes j and m.
$(h_{ij.km})$ = H-matrix of element $h_{ij.km}$.
H = Determinant of matrix (H).
(H) = H-matrix of network.
H_{ij} = Cofactor of determinant element h_{ij}.
i_i = Current flowing from an external source into node i.
i_{ik} = Current flowing from node k to node i through the branch ik.
$i_{ij.km}$ = Component of current i_{ik} due to voltage v_{jm}.
$(i + k, j + m)$ = Addition operator.
k_{ik}'' = Current-branch admittance function for the branch ik.
k'_{jm} = Voltage-branch admittance function for the branch jm.
$k_{ij.km}$ = Branch-pair admittance.
m = Mutual admittance of transformer.
t = Tree on network.
T = Set of trees on network.
v_j = Potential of node j.
v_{im} = Voltage between nodes j and m, with m positive.
$Y_{ij.km}$ = Transfer admittance of network.

Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Percival is with Electric and Musical Industries, Ltd.

(1) INTRODUCTION

An electrical network is an assembly of suitably connected electrical components such as resistors, inductors, capacitors, valves and transformers. The physical network can be represented by a circuit diagram in which each component is denoted by a symbol which is generally a simplified picture of one form of the component itself. The circuit diagram enables the connections of the network to be seen at a glance, while a closer examination reveals the functions of the various parts.

For mathematical purposes the semi-pictorial representation of the network must be replaced by a mathematical representation. This may consist of a diagram, known in topology as a graph, in which the only significant features are the branches, nodes and meshes. Alternatively the representation may be algebraic, e.g. a set of equations, a matrix or a determinant.

It is well known that a linear network comprising 2-terminal elements only can be represented by a graph. With the aid of mathematical trees it is possible to evaluate network admittances directly from the graph. As shown in Reference 1, the quickest way of obtaining the set of trees is to expand the graph as the sum of a number of simpler graphs from which the set of trees can be written down by inspection.

The fact that the graph of a network can be expanded as the sum of a set of simpler graphs is an illustration of the fact that a graph is a mathematical entity, not only in the static sense that it is a representation, but also in the dynamic sense that, like an algebraic expression, it can be operated upon and expanded as a series of terms.

Since an electrical network can be represented either by its graph or by its determinant it must be possible to construct either one of these from the other. Indeed the quickest way of setting up the determinant of a network comprising 2-terminal elements only is to construct the graph first, and then to apply the well-known rule for forming the elements of the determinant from the graph.

In Reference 2 it was shown that 3- or 4-terminal elements can be represented as graphs in terms of an element $h_{ij.km}$ which comprises four nodes and two directed branches. Such elements include transformers, valves and transistors. Hence it is possible to represent any linear network by a graph.

In Reference 2 the graph of a network was utilized solely for the purpose of setting up matrices and determinants. In the ordinary nodal D-determinant the row and column corresponding to the ground node of the network are omitted. If these are included, an enlarged determinant is obtained in which every node in the graph is represented by a row and a column. This determinant was termed an H-determinant. The corresponding H-matrix has the property that the sum of the H-matrices of the elements of the network is equal to the H-matrix of the network as a whole.

To obtain the determinant of a network no equations are required. The procedure is first to set up the graph of the network including a fictitious element f representing the negative of the network admittance it is desired to evaluated. The H-matrix may then be obtained either by adding the H-matrices of the network elements, or by obtaining the matrix elements one at a

time with the aid of a rule given in Reference 2, which is an extension of the well-known rule for networks containing passive elements only. The D-determinant is then obtained by deleting one row and one column which may be those containing the greatest number of network elements. The required network admittance is obtained by equating the D-determinant to zero.

The purpose of the paper is to lay down the basis for a theory of graphs of networks containing 3- or 4-terminal elements, such as valves and transformers. This theory is much more extensive than that relating to networks with 2-terminal elements only, and it appears that there are many practical applications. Within the limits of a single paper it has therefore seemed best to deal with the fundamental aspects only, in order that a clear general picture can be obtained.

The development of the theory is considerably facilitated by the correspondence between the graph of the network and its H-matrix and H-determinant. By simplifying the type of graph employed in Reference 2, and then splitting it into two separate graphs, termed the "current" and "voltage" graphs, it is possible to maintain a 1 : 1 correspondence between every operation on the H-matrix and every operation on the graphs. Thus it is possible to take the derivative of a graph with respect to a network element in the same way as it is possible to take the corresponding derivative of an H-determinant.

In complicated networks the labour of evaluating the D-determinant may be very considerable. It may be reduced, however, by expanding the graph of the network into the sum of simpler graphs and adding the simpler determinants so obtained.

The theorem that the D-determinant of a network containing 2-terminal elements only is equal to the set of trees on the network is extended to include networks with 3- or 4-terminal elements. By including a fictitious element to represent the negative of the required network admittance its value can be obtained merely by equating the set of trees to zero. No 2-trees or linkages, as employed in Reference 1, are required.

(2) THE H-MATRIX AND H-DETERMINANT

For reference purposes it will be necessary to reproduce some of the diagrams and equations of Reference 2. Fig. 1 shows the

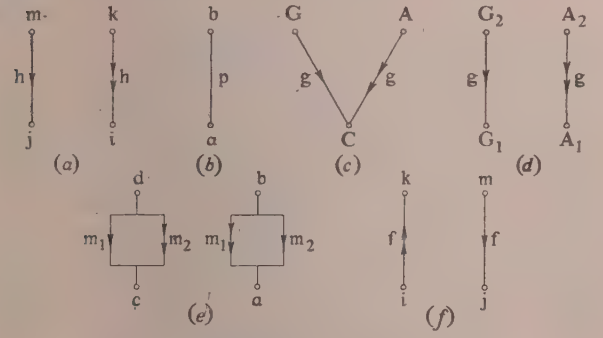


Fig. 1.—Graphs of the network elements.

(a) $h_{ij.km}$
(b) $p_{aa.bb}$
(c) $g_{GC.AG}$
(d) $g_{A1G1.A2G2}$
(e) $(m_{ac.bd}) + (m_{ca.bd})$
(f) $f_{kj.im} = Y_{ij.km}$

graphs of the various network elements. Fig. 1(*a*) shows the fundamental element $h_{ij.km}$, which is such that a voltage v_{jm} between the nodes j and m, with m positive, produces a current i_{ik} from k to i given by

$$h_{ij.km}v_{jm} = i_{ik} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Fig. 1(*b*) shows an ordinary 2-terminal element obtained from $h_{ij.km}$ by identifying i and j and also m and k. Fig. 1(*c*) shows a valve, and Fig. 1(*d*) shows a pair of identical valves in push-pull. Fig. 1(*e*) shows a symmetrical 2-way mutual admittance, which may be that of a transformer. $f_{kj.im}$, shown in Fig. 1(*f*), is a fictitious element equal to a transfer admittance which it is desired to evaluate.

In the case of valves, passive elements such as the input and output admittances are assumed to be included in the external network. A transformer with primary inductance L_1, secondary inductance L_2 and mutual inductance M is replaced by a mutual admittance m with a shunt admittance p_1 across the primary winding and a shunt admittance q_1 across the secondary winding, where

$$1/p_1 = j\omega L_1(1-k^2);\ 1/q_1 = j\omega L_2(1-k^2);\ 1/m = j\omega M(k^2-1) \quad . \quad . \quad . \quad . \quad (2)$$

where ω is 2π times the frequency and $k = M/\sqrt{(L_1L_2)}$ is the coupling coefficient. As in the case of valves p_1 and q_1 are assumed to be included in the external network.

Eqn. (3) gives the H-matrix of the element $h_{ij.km}$ from which the matrices of all other elements can be deduced:

$$(h_{ij.km}) = \begin{array}{c|ccc} & j & m \\ i & h & -h \\ k & -h & h \end{array} \quad . \quad . \quad . \quad . \quad (3)$$

Fig. 2(*a*) is the conventional representation of a network fed with a current i and from which it is required to obtain the

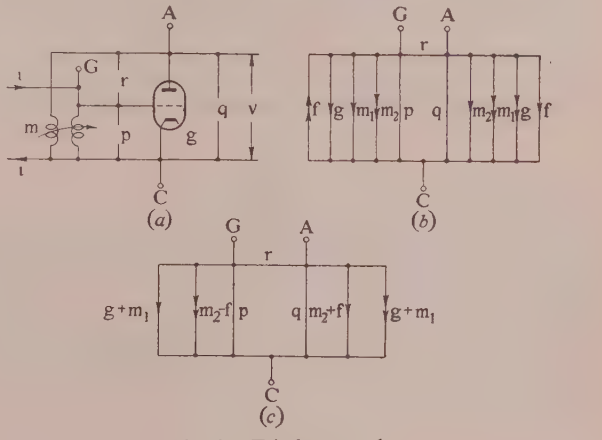


Fig. 2.—Triode network.

output voltage v. The network has been constructed to include a valve of mutual conductance g, a transformer of mutual admittance m and three self-admittances p, q and r. The required transfer admittance is $Y_{CC.GA} = i/v$.

Fig. 2(*b*) shows the graph of this network in which the values of m_1 and m_2 are both m, and $f_{GC.CA} = Y_{CC.GA}$. The current branch f, with the double arrow, carries the current i which is maintained indefinitely by the output voltage v across the voltage branch f, with the single arrow. The effect of this feedback element $f_{GC.CA}$ is to produce a network in dynamic equilibrium. For such a network the determinant $D = 0$, from which $f_{GC.CA}$, and hence $Y_{CC.GA}$, can be calculated.

The H-matrix of Fig. 2(*b*) is

$$(H) = \begin{array}{c|ccc} & C & G & A \\ C & p+q+2m+g-f & -p-m-g & -q-m+f \\ G & -p-m+f & p+r & -r+m-f \\ A & -q-m-g & -r+m+g & q+r \end{array} \quad (4)$$

in which each network element appears in the matrix element h_{ij} in accordance with eqn. (3). Each matrix element can be determined separately by the rule that h_{ij} is equal to the sum of all the network elements whose current branches terminate on node i and the voltage branches on node j, the sign being positive if the arrows on both branches point either towards or away from the nodes i and j, and negative otherwise.

The ordinary nodal D-determinant of a network is equal to any first cofactor of the determinant H of the H-matrix, and for Fig. 2 it is therefore given by

$$\partial H/\partial h_{CC} = H_{CC} = D = pq + pr + qr + rg + 2rm - mg - m^2 + f(m + g - r) \quad . \quad (5)$$

which, since the network is in equilibrium, is zero. Hence

$$Y_{CC.GA} = f_{GC.CA} = \frac{pq + pr + qr + rg + 2rm - mg - m^2}{r - m - g} \quad (6)$$

(3) THE GRAPHS OF A NETWORK

In the graph of Fig. 2(*b*) each network element is shown separately. In this Section it will be shown how such a graph can be reduced to a simpler and more useful form, while other types of graph can be obtained each of which is necessary to the general theory and has its particular applications.

(3.1) The Reduced Graph

In Fig. 2(*b*) it will be seen that both g and m_1 are of the form $h_{CC.AG}$, while $-f$ and m_2 are of the form $h_{CC.GA}$. Such elements are in parallel, and from the point of view of the graph, they constitute a single element. We shall therefore introduce a new symbol k to denote the sum of elements which have both their current and voltage branches and parallel. Thus we have

$$k_{ij.km} = \sum_h h_{ij.km} \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

which gives $k_{CC.AG} = g + m_1$ and $k_{CC.GA} = -f + m_2$.

Fig. 2(*c*) is the graph of the network in terms of the k's, which will be termed *branch-pair admittances*. Thus $k_{CC.AG}$ is the admittance of the branch pair comprising the current branch CA and the voltage branch CG.

The graph of Fig. 2(*c*) can be further simplified by replacing each set of parallel branches by a single branch as in Fig. 3, which will be termed the *reduced graph* of the network.

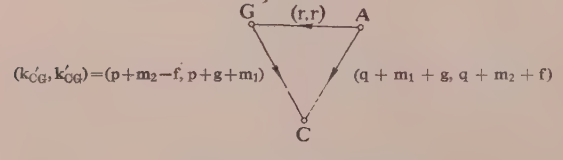


Fig. 3.—The reduced graph.

The first step in the reduction is to cause all arrows on parallel branches to point in the same direction, the signs of the branches being changed where necessary. In the example given no change is required, since the sign of f has already been changed in the process of replacing network elements by branch-pair admittances. Each branch in the reduced graph is then denoted by a pair of terms enclosed in a bracket, the first term being the sum of the current branches and the second the sum of the voltage branches. If the direction of the arrow on the branch is reversed, all the signs in the brackets must be reversed.

In order to maintain the 1 : 1 correspondence between the graph and the H-matrix it is convenient to impose a further condition, namely that the arrows on the branches always point from a higher node to a lower node, a node being higher if it corresponds to a row or column occurring later in the H-matrix.

It will be appreciated that the reduced graph is not only simpler than the previous forms but renders it easier to apply the rule given in the previous Section for obtaining the H-matrix from the graph. As will be shown it is also of greater importance in the theory of graphs.

(3.2) The Current and Voltage Graphs

Each branch on the reduced graph can be considered as being composed of two branches—a composite current branch and a composite voltage branch. Hence the reduced graph can be replaced by two graphs, a *current graph* as in Fig. 4(*a*) composed of current branches only, and a *voltage graph* composed of voltage branches only.

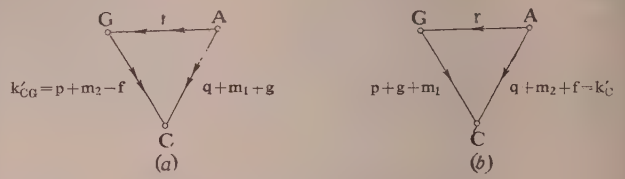


Fig. 4.—Current and voltage graphs.

The total current in each composite current branch is the sum of the currents in the component current branches. Hence if each branch of the current graph is labelled with the total current taken by the branch, the graph completely describes the properties of the network so far as currents are concerned. The only condition which must be imposed on the current graph considered alone is that Kirchhoff's law for currents must be obeyed.

If each node of the voltage graph is labelled with its voltage, this graph completely describes the properties of the network so far as voltages are concerned. The only condition which must be imposed on the voltage graph considered alone is that Kirchhoff's law for voltages must be obeyed.

The current and voltage networks are linked by the admittances of the elements. Thus the element m_2 is associated with the current branch CG and the voltage branch CA. Hence the current flowing from G to C due to this element is m times the voltage of A relative to that of C.

The nodes on the reduced graph correspond both to the rows and columns of the H-matrix. But the nodes on the current graph correspond to the rows, and the nodes on the voltage graph correspond to the columns, of the H-matrix. Hence, as will appear later, the correspondence between the current and voltage graphs and the H-matrix extends even to operations on the determinant of the H-matrix in which the rows are different from the columns. This is a very important advantage, since an essential feature of the theory of graphs is its correspondence with the theory of H-matrices and determinants.

(3.3) The Nodal Graphs

The rule for obtaining the matrix elements from the graph of the network depends on associating each element with a pair of nodes. This association may be made most directly from the nodal graphs of the network.

Fig. 5 shows what will be termed the *reduced nodal graph* in which each node is labelled with the sum of the expressions for the branches terminating on the node, the sign being positive if the arrow on the branch points towards the node and negative if it points away. It will be observed that all information relating

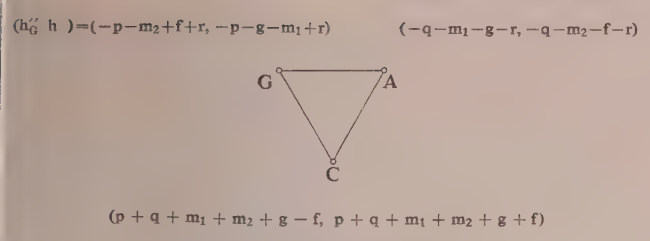


Fig. 5.—Reduced nodal graph.

to the network is retained without the necessity for putting arrows on the branches. In fact there is no need to show the branches at all.

The *current and voltage nodal graphs* are shown in Fig. 6. Since no arrows appear it is necessary to label the current graph i and the voltage graph v.

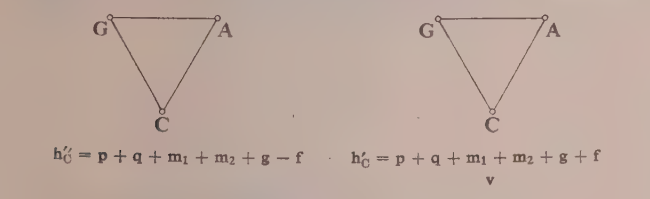


Fig. 6.—Current and voltage nodal graphs.

The matrix element h_{ij} is obtained from the nodal graphs merely by taking the elements common to the current node i and the voltage node j and affixing a positive sign if the sign of the element is the same for both nodes and a negative sign otherwise. This may be checked by obtaining eqn. (4) from Fig. 5 or 6.

The element h_{ij} will also be termed the *node-pair admittance* since, as will be shown in the next Section, it bears the same relation to the nodal graphs as the branch-pair admittance bears to the branch graphs.

(4) THE ALGEBRA OF THE GRAPHS

The algebra of the graphs deals with the relations between the various admittances and admittance functions appearing in the graphs and with the relations between the graphs and the H-matrix. It also deals with the correspondences between operations on the graphs and on the H-matrix, but the treatment of this aspect will be postponed to later Sections.

(4.1) Branch Admittances

In Fig. 4 the component of current flowing from G to C due to the voltage v_{CA} is

$$i_{CC.GA} = (m - f)v_{CA} = k_{CC.GA}v_{CA} \quad . \quad . \quad . \quad (8)$$

while the total current flowing through the branch from G to C is

$$\begin{aligned} i_{CG} &= \sum_{j,m} h_{Cj.Gm}v_{jm} \\ &= (m - f)v_{CA} + pv_{CG} \\ &= k_{CC.GA}v_{CA} + k_{CC.GG}v_{CG} \quad . \quad . \quad . \quad . \quad (9) \end{aligned}$$

so that $k_{CC.GA} = \partial i_{CG}/\partial v_{CA}$ and, in general,

$$k_{ij.km} = \sum_{h} h_{ij.km} = \partial i_{ik}/\partial v_{jm} \quad . \quad . \quad . \quad (10)$$

in which it is necessary to interpret the partial derivative as implying that all voltages other than v_{jm} are ineffective in producing currents. The interpretation that all voltages other than v_{jm} are zero is not possible, as voltages round a mesh are not independent.

The expression which has been used as a label for a current branch in the current graph will be termed the *current-branch admittance function* and written k''_{ik} for the current branch ik with the arrow pointing towards i. Thus $k''_{CG} = p + m_2 - f$. The expression which has been used as a label for a voltage branch in the voltage graph will be termed the *voltage-branch admittance function* and written k'_{jm} for the voltage branch jm with the arrow pointing towards j. Thus $k'_{CA} = q + m_2 + f$.

The expression in the brackets used in the single reduced graph as a label for a branch can therefore be written (k''_{ik}, k'_{ik}). Thus the branch CG in Fig. 3 would appear as $(k''_{CG}, k'_{CG}) = (p + m_2 - f, p + g + m_1)$.

$k_{ij.km}$ is the sum of the elements common to k''_{ik} and k'_{jm}, the sign being positive for an element which has the same sign in k''_{ik} and k'_{jm}, and negative for an element which has the opposite sign. This relationship can be expressed in the form

$$k_{ij.km} = k''_{ik} \,.\, k'_{jm} \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

in which the dot indicates a type of multiplication in which $A.A = A$ and $A.B = 0$, the signs being combined by the ordinary rule for multiplication. Thus in Fig. 4

$$k_{CC.GA} = k''_{CG} \,.\, k'_{CA} = (p + m_2 - f) \,.\, (q + m_2 + f) = m_2 - f \quad . \quad . \quad . \quad . \quad . \quad (12)$$

We also have

$$k''_{ik} = \sum_{j,m} k_{ij.km}; \quad k'_{jm} = \sum_{i,k} k_{ij.km} \quad . \quad . \quad . \quad (13)$$

bearing in mind that a reversal of any pair of letters denoting a branch reverses the sign.

(4.2) Node Admittances

The expression which has been used as a label for a node in the current nodal graph will be termed the *current-node admittance function* and written h''_i for the node i. Thus $h''_G = -p - m_2 + f + r$. The expression which has been used as a label for a node in the voltage nodal graph will be termed the *voltage-node admittance function* and written h'_j for the node j. Thus $h'_A = -q - m_2 - f - r$.

The expression in the brackets used in the single nodal graph as a label for a node can therefore be written (h''_i, h'_j). Thus the node G in Fig. 5 is

$$(h''_G, h'_G) = (-p - m_2 + f + r, -p - g - m_1 + r)$$

From the way in which the nodal admittance functions were derived in Section 3.3, we have

$$h''_i = \sum_{k} k''_{ik}; \quad h' = \sum_{m} k'_{jm} \quad . \quad . \quad . \quad . \quad (14)$$

again remembering that a reversal of any pair of letters denoting a branch reverses the sign.

The rule given in Section 3.3 for forming the matrix element, or node-pair admittance, h_{ij}, leads to the equation

$$h_{ij} = h''_i \,.\, h'_j \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

in which the dot has exactly the same significance as in eqn. (11) for the branch-pair admittance. Thus

$$h_{GA} = (-p - m_2 + f + r) \,.\, (-q - m_2 - f - r) = m - f - r \quad . \quad . \quad . \quad . \quad (16)$$

in accordance with eqn. (4).

(4.3) Derivation of the *H*-Matrix

In Reference 2 it was shown that the nodal equation for the *i*th node of a network can be written in the form

$$\sum_j \sum_{k,m} (h_{ij.km} - h'_{kj.im} - h''_{im.kj} + h'''_{km.ij})v_j = i_i \quad . \quad (17)$$

where i_i is the current flowing from an external source into node *i*. For a network in equilibrium $i_i = 0$.

It is now possible to write this equation in the simpler form

$$\sum_j \sum_{k,m} k_{ij.km} v_j = i_i \quad . \quad . \quad . \quad . \quad . \quad (18)$$

The quantity h_{ij} is defined by the equation

$$h_{ij} = \sum_{k,m} k_{ij.km} \quad . \quad . \quad . \quad . \quad . \quad (19)$$

The complete set of equations, one for each node, can then be written

$$\sum_j h_{ij} v_j = i_i \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

which represents an equation for each value of the suffix *i*. As before, the right-hand side of all these equations is zero for a network in equilibrium.

From eqn. (20) it follows that

$$h_{ij} = \partial i_i / \partial v_j \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

in which the voltages of all the nodes except the *j*th are held constant.

By comparing eqns. (11) and (15) it will be seen that the relation between branch-pair admittances and the branch admittance functions is similar to the relation between node-pair admittances and node admittance functions. By comparing eqns. (10) and (21) it will be seen that the relation between branch-pair admittances and the branch currents and voltages is similar to the relation between node-pair admittance and the nodal currents and voltages.

The rule given in Section 2 for deriving the matrix elements from the network elements is equivalent to the first part of eqn. (10) together with eqns. (13), (14) and (15). By employing the branch-pair admittance graph of Fig. 2(*c*), eqn. (10) is no longer required, and by employing the reduced graph of Fig. 3, eqn. (13) need not be used, while by utilizing the reduced nodal graph of Fig. 5 only eqn. (13) is necessary.

In practical work with the graphs and *H*-matrices it has been found most convenient to set up the *H*-matrix from either the reduced branch graph or the current- and voltage-branch graphs.

(5) CORRESPONDING OPERATIONS ON *H*-DETERMINANTS AND GRAPHS

(5.1) The Operator $\partial H / \partial h_{ij.km}$

It was shown in Reference 2 that the derivative of an *H*-determinant with respect to the network element $h_{ij.km}$ is given by

$$\partial H / \partial h_{ij.km} = (i + k, j + m)H \quad . \quad . \quad . \quad (22)$$

where $(i + k, j + m)$ is an addition operator which deletes row *k*, replaces row *i* by the sum of the rows *i* and *k*, performs the corresponding operations on the columns *j* and *m* and affixes the sign $(-1)^{k+m}$. A derivative may be taken with respect to an element which does not appear in *H*, since it is only necessary to state that such a derivative is given by the corresponding addition operation.

By comparing the *H*-determinant with its graphs it can be seen that the operation of adding rows in the *H*-determinant corresponds to short-circuiting the corresponding nodes in the current graph, while the operation of adding columns corresponds to the operation of short-circuiting the corresponding nodes in the voltage graph. Affixing the sign $(-1)^{k+m}$ to the *H*-determinant corresponds to affixing the sign $(-1)^k$ to the current graph and the sign $(-1)^m$ to the voltage graph.

An operation on a graph corresponds to an operation on the *H*-determinant if the resulting determinant is the *H*-determinant of the resulting graph. It was shown in Reference 2 that the derivative of an *H*-determinant with respect to a network element is also an *H*-determinant. If the sign of the *H*-determinant of a pair of current and voltage graphs is defined as the product of the signs of the graphs it can easily be proved that the *H*-determinant of the graph formed in the manner described above is the derivative of the *H*-determinant of the original graph.

If *H* in eqn. (22) represents the current and voltage graphs and $i + k$ denotes the operation of short-circuiting the nodes *i* and *k* in the current graph and affixing the sign $(-1)^k$, and $j + m$ denotes the operation of short-circuiting the nodes *j* and *m* in the voltage graph and affixing the sign $(-1)^m$, it follows that the resulting graphs are the derivatives of the original graphs with respect to the network element $h_{ij.km}$.

The derivative of the *H*-determinant of the network of Fig. 2 with respect to the element *g* is given by

$$\partial H / \partial g = (C+A, C+G)H = -\begin{matrix} C+G & A \\ C+A & -r+m-f \\ G & r-m+f \end{matrix} \quad \begin{matrix} \\ r-m+f \\ -r+m-f \end{matrix} \quad . \quad . \quad . \quad . \quad . \quad (23)$$

The corresponding current and voltage graphs are derived from Fig. 4 and consist of single branches as shown in Fig. 7,

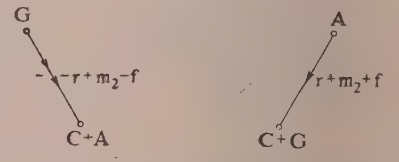


Fig. 7.—$\partial H / \partial g$ = Derivative of the graphs of Fig. 4 with respect to $g_{CC.AG}$.

a negative sign being affixed to the current graph and a positive sign to the voltage graph. The elements *g* and m_1, which are in parallel in both graphs, are short-circuited. The branch of the element *q* and the voltage branch of the element *p* are also short-circuited. Hence these elements must be deleted. This is in accordance with the rule, given in Reference 2, that, in taking the derivative with respect to a given network element, all elements which are in either current or voltage parallel with it are eliminated.

The nodes formed by short-circuiting two nodes are denoted by the sum of the letters denoting these nodes, the order being the same as in the operator.

It is possible to continue to take derivatives of an *H*-determinant with respect to successive network elements until only a single row and a single column remain with a single determinant element of zero. This can be illustrated from eqn. (23) by taking a further derivative with respect to *r*, the operator being $(G + C + A, C + G + A)$. The corresponding operations on the current and voltage graphs leave only a single node in each graph, the determinant being that described above.

In taking the derivative with respect to an ordinary element $h_{aa.bb}$ the nodes to be short-circuited are the same in the current and voltage graphs. The operation can therefore be applied to a single reduced graph such as that of Fig. 3. The mathematical operation of taking a derivative then corresponds to the electrical

operation of short-circuiting two nodes on the network. It is also equivalent to making the admittance of the element $h_{aa.bb}$ infinite.

It can easily be shown that the mathematical operation of taking the derivative with respect to any element $h_{ij.km}$ is equivalent to making the mutual admittance of this element infinite. Thus eqn. (23) and Fig. 7 represent the result of making the slope g of the valve infinite. The D-determinant of the H-determinant of eqn. (23) is any first cofactor of H and is therefore, on taking the sign into account, $r - m + f$. This is the coefficient of g in eqn. (5), as it should be. But it is also the D-determinant when g is infinite. It follows that as g is increased indefinitely the transfer admittance approaches the value $Y_{CC.GA} = m - r$.

(5.2) The Operator $h_{ij.km}\partial/\partial h_{ij.km}$

The operator $h_{ij.km}\partial/\partial h_{ij.km}$ adds all the network elements in row k, except $h_{ij.km}$, to row i, and all the network elements in column m, except $h_{ij.km}$, to column j. The number of rows and columns is unchanged and no question of sign arises. The D-determinant of the resulting H-determinant gives the set of terms containing $h_{ij.km}$ in the D-determinant of the original network.

From eqn. (4) we have

$$(g_{CC.AG}\partial/\partial g_{CC.AG})H = \begin{array}{c|ccc} & C & G & A \\ C & -r+m-f+g & -g & r-m+f \\ G & r-m+f & 0 & -r+m-f \\ A & -g & g & 0 \end{array} \quad . \quad . \quad . \quad . \quad (24)$$

The D-determinant is $g(r - m + f)$, which is the set of terms containing g in eqn. (5).

The operation $h_{ij.km}\partial/\partial h_{ij.km}$ on the current and voltage graphs is a transfer operation, which leaves $h_{ij.km}$ unchanged, but transfers to node i the ends of all other current branches which are normally connected to node k, and transfers to node j the ends of all voltage branches which are normally connected to node m.

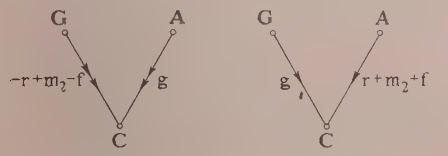


Fig. 8.—Result of operation $g_{CC.AG}\partial/\partial g_{CC.AG}$ on the graphs of Fig. 4.

Fig. 8 shows the effect of the operation $g_{CC.AG}\partial/\partial g_{CC.AG}$ on the graphs of Fig. 4.

(5.3) The Operator $(1 + h_{ij.km}\partial/\partial h_{ij.km})$

H can be expanded in terms of the network element $h_{ij.km}$ by the equation

$$H = H_0 + (h_{ij.km}\partial/\partial h_{ij.km})H \quad . \quad . \quad . \quad (25)$$

where $H_0 = H(h_{ij.km} = 0)$.

But as it is possible to take the derivative with respect to an element which is not actually present in H_0, a more convenient form is

$$H = (1 + h_{ij.km}\partial/\partial h)H_0 \quad . \quad . \quad . \quad . \quad (26)$$

in which the superfluous suffix of h is omitted.

By taking first cofactors we obtain the expansion of D in terms of $h_{ij.km}$ as follows:

$$D = D_0 + (h_{ij.km}\partial/\partial h)D \quad . \quad . \quad . \quad . \quad (27)$$

where $D_0 = D(h_{ij.km} = 0)$.

Corresponding to these expansions of the determinants we have the expansions of the current and voltage graphs. Fig. 9 shows the expansion of the graphs of Fig. 4 in terms of g.

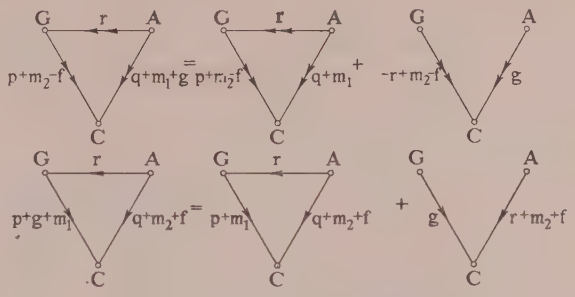


Fig. 9.—$H = H_0 + g_{CC.AG}\partial H/\partial g_{CC.AG}$.

It is possible to continue the expansion in terms of other elements. The resulting expansion of the current and voltage graphs corresponds to the network equations described in Reference 1, but generalized to include all elements of the form $h_{ij.km}$.

(6) EQUIVALENCE OF THE D-DETERMINANT TO THE SET OF TREES ON A NETWORK

It is known that the nodal D-determinant of a network, all the elements of which are of the form $h_{ii.jj}$, is equal to the set of trees on the graph of the network. It will now be shown that, with an appropriate definition of a tree, the theorem can be extended to networks containing elements of the general form $h_{ij.km}$.

(6.1) Trees on the Current and Voltage Graphs

On a graph, all the elements of which are of the form $h_{ii.jj}$, a tree is defined as a set of branches which connect all the nodes but enclose no meshes. It is easily shown that on such a graph any tree can be formed by a succession of transfer operations of the form $h_{ii.jj}\partial/\partial h$. Moreover, on the current and voltage graphs of any network the more general operations $h_{ij.km}\partial/\partial h$, if continued until no meshes or parallel branches remain, give rise to a pair of trees, one on the current graph and the other on the voltage graph.

Thus the graphs of Fig. 8 represent the first step in the formation of a pair of trees. A second and final step consists in the

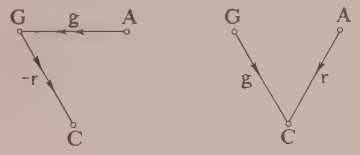


Fig. 10.—The tree rg.

operation $r_{GC.CA}\partial/\partial r$, which gives the pair of trees shown in Fig. 10. The H-determinant of the pair of trees is given by

$$(r_{GC.CA}\partial/\partial r)(g_{CC.AG}\partial/\partial g)H = \begin{array}{c|ccc} & C & G & A \\ C & -r & 0 & r \\ G & r+g & -g & -r \\ A & -g & g & 0 \end{array} \quad . \quad (28)$$

It has already been shown that the operation $h_{ij.km}\partial/\partial h$ gives an H-determinant, the D-determinant of which is the set of terms in the D-determinant of the original network which contains $h_{ij.km}$. Hence a succession of operations of this type

continued until no further elements remain must give a single term in D. Thus the D-determinant of eqn. (28) gives the single term rg of eqn. (5).

We define a tree on the graphs of a network as a set of network elements such that the current branches form a tree on the current graph, and the voltage branches form a tree on the voltage graph. Thus the set of elements r and g form a tree on the graphs of Fig. 4. Another tree is formed by the elements f and m_1. Indeed, as will be proved later, every term in D of eqn. (5) corresponds to a tree on the graphs of Fig. 4.

In the case of a network composed entirely of elements of the form $h_{ii.jj}$ the current and voltage graphs are identical, and the definition of a tree given above reduces to the standard definition. In the more general case the graphs are not identical and neither are the trees on the graphs. Thus if a tree includes the element $h_{ij.km}$ the tree on the current graph will include the branch ik, while the corresponding branch on the tree on the voltage graph will be jm.

For a network composed entirely of elements of the form $h_{ii.jj}$ the value t of a tree on the graph of the network is defined as the product of its elements. This is equal to the D-determinant of the tree, which in turn is equal to one of the terms in the D-determinant of the network. The set T of trees on the network is then defined as the sum of the individual trees, from which follows the theorem that the set of trees on a network is equal to the nodal D-determinant of the network,

i.e. $$T = \Sigma t = D \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

It is required to extend this theorem to the general case of a network composed of elements of the type $h_{ij.km}$. We commence by defining the value t of a tree, not directly in terms of the graphs, but as the D-determinant of the graphs of the tree.

At this point it is necessary to draw attention to the fact that the tree rg of Fig. 10 is not identical with the tree rg on the graphs of Fig. 4. Nevertheless it can be shown that the value of the D-determinant of the tree of Fig. 10 is the same as the value of the D-determinant of the tree rg on the graphs of Fig. 4. Hence the values of the trees are the same. The tree rg of Fig. 4 can be isolated by the operation $rg(C + A, C + G)(A + G, G + A)$, the first operation being $(A + G, G + A)$, which can be shown to be equivalent to the operation of eqn. (28). It can be shown that operations always exist whereby a tree can be isolated in its original form, but that, in general, other operations exist whereby a tree can be formed having the same value but with its branches connected between different nodes.

It thus follows that D is the sum of terms each of which is given by a succession of operations of the type $h_{ij.km}\partial/\partial h$, and that each tree is also given by a succession of operations of this type. Hence the value of each tree is equal to one of the terms in D, and D itself is equal to the set of trees T. Hence eqn. (29) is true in the general case.

It remains to define the value of a tree in terms of the graphs of the network, so that it is unnecessary to refer to determinants in order to ascertain the value of the set of trees, and then to give a rule whereby T can be obtained without enumerating the individual trees.

(6.2) The Value of a Tree

For a network composed entirely of elements of the form $h_{ii.jj}$ the determinant of a tree is simply the product of the elements of the tree. In the general case the determinant is equal to the product of the elements, but with a sign which must be determined. If the determinant is expanded by means of addition operators it may be shown that the process is equivalent to the following definition for the value of the equivalent tree:

Multiply the branches in each graph together and, provisionally, give the tree a positive sign if the signs of the products are the same, and a negative sign otherwise. In each graph, associate each branch with a node. If the arrow on the branch points away from the node write down a negative sign. If the total number of negative signs is odd change the sign of the tree; otherwise make no change. Finally, assign a convenient order to the nodes and write down the branches of the tree on each graph in the order of the associated nodes, including, for each graph, a dummy branch to correspond with the node which, it will be found, is not included. If the relative inversions of the branches are odd, change the sign of the tree; otherwise make no change.

Thus for the tree rg of Fig. 4 we write $r(G), g(C), d(A)$ for the current graph and $r(G), g(C), d(A)$ for the voltage graph, where d is the dummy branch. The value of the tree is therefore rg. For the tree m_1m_2 we write $m_1(C), -m_2(G), d(A)$ for the current graph and $m_1(C), -d(G) - m_2(A)$, for the voltage graph. The value of the tree is therefore $-m_1m_2 = -m^2$, corresponding to the term in eqn. (5).

It will be appreciated that, owing to the need for ascertaining the sign, the process of enumerating the individual trees is an even more tedious process than for a passive network.

(6.3) Evaluation of a Set of Trees

In Reference 1 it was shown that, with the aid of network equations, the set of trees on a network could be evaluated more rapidly than by the enumeration of the individual trees. This is also the case for networks containing elements of the more general type $h_{ij.km}$. However, the expansion should always be in terms of ordinary elements.

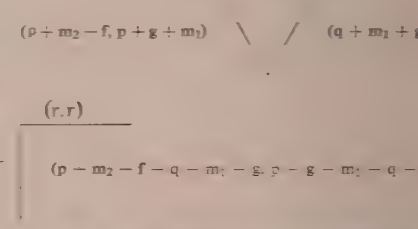


Fig. 11.—Expansion of the graphs of Fig. 3.

Thus Fig. 11 shows the expansion of the graph of Fig. 3 in terms of the ordinary element r, which gives a pair of trees in which, however, the branches are composite. Since the corresponding current and voltage graphs are identical there is no need to show the arrows or to label the nodes, the sign of a term being determined merely by the signs of the branches and the number of relative inversions. Formally we write

$$T = (p + m_2 - f)(q + m_1 + g) \,.\, (p + g + m_1)(q + m_2 + f) + r(p + m_2 - f + q + m_1 + g) \,.\, r(p + g + m_1 + q + m_2 + f) \quad . \quad . \quad . \quad . \quad (30)$$

in which the dot indicates a form of multiplication in which, if A and B are the two factors to be multiplied, $A \,.\, B = 0$ and $A \,.\, A = \pm 1$, the sign being positive if the relative inversions of the individual factors are even, and negative if they are odd. Thus we have for one of the terms in the first dot product $m_2g \,.\, gm_2 = -mg$, and for the complete expression

$$T = D = rp + 2rm - rf + rq + pq - m^2 - mg + mf + gf + rg \quad . \quad . \quad . \quad . \quad (31)$$

which is the same as in eqn. (5). It will be noted that the suffices of m must be retained in eqn. (30) but are not required in eqn. (31).

(6.4) An Example

The following example is of some intrinsic interest, and shows how the tree method can be used to advantage.

Wheeler[3] has introduced the concept of an ideal transformer-repeater which has the following properties:

(*a*) It has two pairs of terminals which will be referred to as A and B.
(*b*) If A (or B) is on open-circuit the admittance of B (or A) is zero.
(*c*) If A (or B) is on closed circuit the impedance of B (or A) is zero.
(*d*) The current multiplication from A to B is a and that from B to A is $1/a$, where a may be complex.
(*e*) The voltage multiplication from A to B is b and that from B to A is $1/b$, where b may be complex.

It follows that on open- or closed-circuit no power is taken by the device from a source at either A or B, that the impedance transformation from A to B is b/a and that from B to A is a/b, and that the power gain from A to B is ab and that from B to A is $1/ab$.

It will be shown that this concept can be realized as a limiting case, with appropriate limitations on the upper frequency, by means of the network shown in Fig. 12(*a*), in which p and q

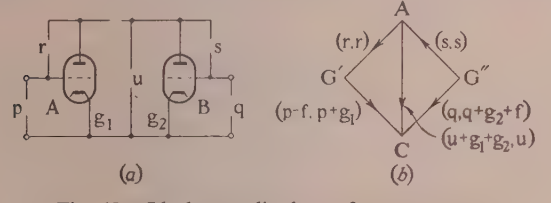


Fig. 12.—Ideal generalized transformer-repeater.

are external admittances, r and s are small admittances, u is a coupling admittance and g_1 and g_2 are large mutual admittances. The valves shown may each be replaced by several valves in tandem whereby the values of g_1 and g_2 may be made complex and as large as required.

The reduced graph is shown in Fig. 12(*b*), in which $f_{G'C.CG''}$ is the transfer admittance from A to B. This can be expanded in terms of r and s and solved as in the previous example. However, the graph can be considerably simplified by taking advantage of the fact that g_1 and g_2 are very large.

It follows that all the terms in the determinant can be neglected with the exception of those containing g_1 or g_2. Hence in the graph the only trees which need to be considered are those containing g_1 or g_2. But the only current branch containing g_1 or g_2 is $u + g_1 + g_2$, in which u can be neglected, which leaves $g_1 + g_2$. It follows that the element u does not appear in the determinant, and hence the voltage branch u can be deleted.

We now expand the network in terms of g_1 and g_2, as explained in Section 5.3. The end A of the current branch r must be transferred to C so that the current branch r becomes in parallel with the current branch $p - f$, but with its sign reversed owing to the opposite direction of the arrows. Similarly the current branch s appears in parallel with the current branch q with the same sign. If the voltage part of the graph is expanded in terms of $p + g_1$, the branch r takes the place of the voltage branch u with the same sign. The voltage branch $q + g_2 + f$ will be unchanged. The voltage branch s can also be left unchanged.

We have now obtained a current tree and a voltage graph. In order to make these identical it will be convenient to expand the latter in terms of the branch $q + g_2 + f$. This is quite a permissible operation although its only effect is to transfer the end G'' of the voltage branch s to C, when the branch becomes in parallel with r but of opposite sign. This gives the single tree of Fig. 13.

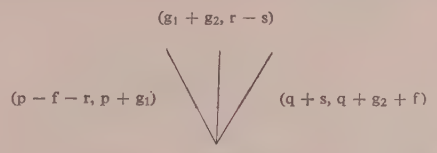


Fig. 13.—Graph of Fig. 12(*b*) when g_1 and g_2 are large.

The value of this tree is

$$T = D = (p - f - r)(g_1 + g_2)(q + s).$$
$$(p + g_1)(r - s)(q + g_2 + f)$$
$$= pg_2s + rg_1q + fg_1s \quad . \quad (32)$$

By putting $D = 0$ and $p = 0$ we obtain the transfer admittance

$$f = Y_{CC.G'G''} = -qr/s \quad . \quad . \quad . \quad . \quad (33)$$

and by symmetry

$$Y_{CC.G''G'} = -ps/r \quad . \quad . \quad . \quad . \quad (34)$$

It is easily seen that by deleting f and replacing p by $-f'$ we obtain the input admittance at A

$$Y_{CC.G'G'} = qrg_1/sg_2 \quad . \quad . \quad . \quad . \quad . \quad (35)$$

which is infinite when $1/q = 0$ and zero when $q = 0$.

From eqn. (33) the current gain is $a = -s/r$, while from eqns. (33) and (35) the voltage gain is $b = -g_1/g_2$. The conditions for the ideal transformer-repeater are therefore satisfied.

(7) PRACTICAL APPLICATIONS

In dealing with the graphs of networks it has been necessary to lay the main stress on the theoretical relations between the various aspects. A brief summary will now be given of the practical applications of the results obtained in this and the two preceding papers.[1,2]

To solve simple networks consisting of 2-terminal elements only, the choice of method is of little importance and the usual D-determinant method is as good as any other. If the network is more complicated the H-determinant method is generally quicker, since the row and column deleted to give the D-determinant may be chosen as those containing the greatest number of network elements. The H-determinant method also has an advantage when the transfer admittance involves four terminals, since only a single cofactor is required.

The advantage of determinant methods as compared with tree methods is that fewer rules are required. If the labour involved in expanding the determinant is considerable the tree method is quicker, since there are no terms to be cancelled in the D-determinant.

For networks containing valves or transformers it has been found of considerable advantage to construct the reduced graph of the network as shown in Fig. 3. Either the D-determinant or the H-determinant can be set up from the graph, but the latter has generally been found most useful. If a transfer admittance is required, the fictitious element f may be included in the graph when it becomes unnecessary to take cofactors. The network admittance across any 2-terminal element p can always be found by replacing p by $p - f'$ in the D-determinant and equating it to zero.

If valves or transformers are present the tree method is inevitably more complicated and is to be recommended only in special cases such as the example given. However, the labour of expanding the determinant of a very complicated network may be

considerably reduced by expanding the graph as the sum of simpler graphs and then adding the determinants of these graphs. Equivalent networks can be found with the aid of *H*-matrices as described in Reference 2. It has also been found that useful equivalent networks can be obtained directly from the graphs.

(8) ACKNOWLEDGMENTS

The author is indebted to E.M.I. Research Laboratories, Ltd., for permission to publish the paper and to colleagues on the staff for valuable discussions.

(9) REFERENCES

(1) PERCIVAL, W. S.: "The Solution of Passive Electrical Networks by Means of Mathematical Trees," *Proceedings I.E.E.*, Paper No. 1492 R, May, 1953 (**100**, Part III, p. 143).

(2) PERCIVAL, W. S.: "Improved Matrix and Determinant Methods for Solving Networks," *Proceedings I.E.E.*, Monograph No. 96 R, April, 1954 (**101**, Part IV, p. 258).

(3) WHEELER, H. A.: Wheeler Monographs, No. 5, Vol. I, Wheeler Laboratories.

621.317.331

Monograph No. 130 M
Apr. 1955

RESIDUAL EFFECTS IN THE CAMPBELL BRIDGE METHOD FOR THE ABSOLUTE MEASUREMENT OF RESISTANCE, AND A NOTE ON A NEW BRIDGE

By Professor N. F. ASTBURY, M.A., Sc.D., F.Inst.P., Member.

(*The paper was received 19th October*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in April*, 1955.)

SUMMARY

Systematic errors, associated with coil reversals, in the determination of the absolute unit of resistance by Albert Campbell's a.c. bridge method, originally noted by Hartshorn and Astbury and recently confirmed by Rayner, can, it is shown, be explained by uncompensated eddy-current coupling between supply and detector circuits. It is also shown that the errors are largely, but not necessarily completely, eliminated when a mean value is taken. A new bridge is proposed in which, by measuring an additional mutual inductance, the effect can be eliminated without external compensation: it is also indicated how this might be done, less conveniently, on the Campbell bridge. A model of the new bridge, operating at 1 kc/s, has been used to demonstrate experimentally the introduction of systematic variations associated with coil reversals, the eddy-currents being produced under control in a tertiary circuit.

(1) INTRODUCTION

The recent publication by G. H. Rayner[1] of a new series of measurements at the National Physical Laboratory on the absolute determination of the resistance unit by Albert Campbell's a.c. bridge method has persuaded the present author of the necessity for a reassessment of some of the residual effects in this and similar methods.

The technical problem involved in the absolute determination of resistance is, of course, the comparison of the measure of a physical model of the unit of resistance with the quotient of inductance and time. The inductance is usually that of a coil system of closely-defined geometry and of high stability, and is computed from the dimensions of the system: time enters as a frequency or rate of change, and is referred to the mean solar second. Circuit methods, as distinct from generator methods (of which the Lorenz method[2] is the prime example), seem to fall into two classes, namely those in which a single inductor is used together with some commutation device, and those in which two inductors are used with pure alternating current. The first group includes the work of Grüneisen and Giebe,[3] Curtis, Moon and Sparks,[4] and Thomas, Peterson, Cooter and Kotter,[5] and the second group that of Hartshorn and Astbury,[6] and of Jouaust, Picard and Hérou.[7] Now the main point to be argued in the present paper is concerned with the significance of eddy-current effects arising from the penetration of the magnetic fields of the inductors into neighbouring conducting masses. In the first group of methods such effects can be correctly allowed for by measurements of the frequency dependence of the characteristics of the inductor under the appropriate conditions. In the second group, they were apparently not discussed by Jouaust, Picard and Hérou, but Hartshorn and Astbury recognized the problem and attempted to solve it. This attempt, it will now be argued, was largely ineffective, although the statement of the problem was correct. The essence of the matter is that a mutual impedance is established between supply and detector meshes by eddy-current circuits: reduction of the mutual inductance between these meshes by mechanical adjustments or otherwise is of little significance. We shall now examine the results of Hartshorn and Astbury and of Rayner from this point of view. We shall also propose a new bridge arrangement—a model of which has been used to test the arguments—which, it is suggested, is well worth examination as a precision circuit.

Correspondence on Monographs is invited for consideration with a view to publication.
Professor Astbury is at University College of Khartoum.

(2) SYSTEMATIC ERRORS IN THE N.P.L. CAMPBELL BRIDGE

The Campbell circuit as set up at the N.P.L. is shown in Fig. 1. The equations of balance, including normal residual effects, are

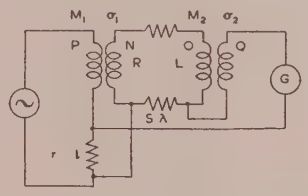


Fig. 1.—N.P.L. Campbell bridge.

set out in Table 1, which also gives the possible signs of the quantities involved. In the circuit as arranged, only the groups with r negative are used, and for these groups the possible connections to the four coils P, N, O, Q (Fig. 1) are set out in Table 2.

Table 1

EQUATIONS OF THE CAMPBELL CIRCUIT AND THE SIGNS INVOLVED

ω^2	$[M_1(M_2$	$+\lambda)$	$+Ll]$	$= Rr$	$+s\sigma_1$	$M_1 s +$	$Lr = 0$
+	+	−	+ +	+ +	− +	+ −	+ +
−	−	+	+ +	+ +	+ +	− +	+ +
+	−	+	+ −	+ −	+ +	+ +	+ −
−	+	−	+ −	+ −	− +	− −	+ −

Table 2

Arrangement	P	N	M_1	O	Q	M_2
1	+	+	+	+	+	−
2	−	−	+	+	+	−
3	−	−	+	−	−	−
4	+	+	+	−	−	−
5	+	−	−	+	−	+
6	−	+	−	+	−	+
7	−	+	−	−	+	+
8	+	−	−	−	+	+

If all the corrections have been successfully estimated and applied, the variations within a group of measurements corresponding to the eight possible arrangements should be small and random. Actually, it was found by Hartshorn and Astbury[6]

and confirmed by Rayner[1] that there were appreciable systematic differences between the individual results for each group, although the variations in the group means was small and apparently random.

It will now be shown that such systematic variations can arise from uncompensated eddy-current effects. In the N.P.L. circuit a so-called zero-adjusting mesh, providing a sort of controlled eddy-current coupling between supply and detector branches, was added to the system. For our present argument this and the actual eddy-current circuits existing in neighbouring conducting masses can be assumed to be integrated into the mesh marked 4 in Fig. 2(*a*). Let μ_{12}, etc., represent total mutual impedances

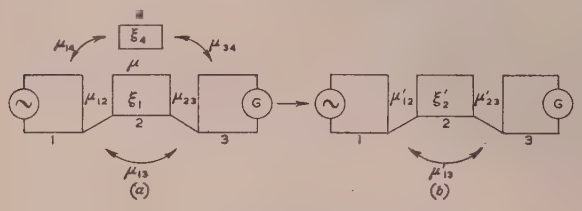


Fig. 2.—Eddy-current transformation.

between the various pairs of meshes, and ξ_2, ξ_4 the self-impedances of meshes 2 and 4 respectively. Since we may reasonably regard the whole system as linear, the 4-mesh system of Fig. 2(*a*) is equivalent to the 3-mesh system of Fig. 2(*b*), in which

$$\mu'_{12} = \mu_{12} - \mu_{14}\mu_{24}/\xi_4 \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$\mu'_{23} = \mu_{23} - \mu_{24}\mu_{34}/\xi_4 \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$\mu'_{13} = \mu_{13} - \mu_{14}\mu_{34}/\xi_4 \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$\xi'_2 = \xi_2 - \mu^2_{24}/\xi_4 \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The balance condition for the system of Fig. 2(*b*) is

$$\mu'_{13}\xi'_2 + \mu'_{23}\mu'_{12} = 0 \quad . \quad . \quad . \quad . \quad . \quad (5)$$

It was thought that the operation of the zero-adjusting mesh secured the condition

$$\mu'_{13} \equiv \mu_{13} \equiv (-)(r + j\omega l)$$

(j is the 90° operator). If, in fact, this condition is not secured, the whole situation is radically changed. As far as ξ'_2, μ'_{12}, μ'_{23} are concerned, the sequence of ancillary measurements determines these without sensible error. Replacing the complex μ's by their explicit form ($\sigma + j\omega m$), σ being impurity (resistive component) and ωm the reactive component, eqn. (5) becomes

$$r + j\omega l + \sigma_{13} + j\omega m_{13} - (\sigma_{14} + j\omega m_{14})(\sigma_{34} + j\omega m_{34})/(r_4 + j\omega l_4) = R' + jX' \quad . \quad (6)$$

where R' and X' are the resistive and reactive components of $(-)\mu'_{23}\mu'_{12}/\xi'_2$, the explicit form of which we do not need for the present. The original results were evaluated on the assumption that

$$r + j\omega l = R' + jX' \quad . \quad . \quad . \quad . \quad . \quad (7)$$

but it appears that we must in fact describe the properties of the resistor r by effective values, r' and l', derived from eqn. (6). Making the plausible assumption that, at the frequencies involved (50c/s and 100c/s) the reactance of the eddy-current paths is negligible compared with their resistance, we can write

$$\left.\begin{aligned} r' &= r + \sigma_{13} - (\sigma_{14}\sigma_{34} - \omega^2 m_{14}m_{34})/r_4 \\ l' &= l + m_{13} - (\sigma_{14}m_{34} + \sigma_{34}m_{14})/r_4 \end{aligned}\right\} \quad . \quad . \quad (8)$$

Returning now to the main equation of balance given in Table 1, we can, without affecting our discussion, omit s and λ, and write

$$\omega^2 M_1 M_2 = Rr - \omega^2 Ll$$

and in the small term $\omega^2 Ll$ we can make the substitution

$$\omega^2 = Rr/M_1M_2$$

giving

$$\omega^2 M_1 M_2 = Rr(1 - Ll/M_1M_2) \quad . \quad . \quad . \quad . \quad (9)$$

From Table 2, it is evident that L/M_1M_2 is a constant which does not change sign when we alter the connections to the coils: let us call it $-p$. Then, replacing r and l in eqn. (9) by r' and l' from eqn. (8), we find, for the essential balance condition, if we retain just the first-order terms,

$$\begin{aligned} \omega^2 M_1 M_2/R = r[1 &+ \sigma_{13}/r - (\sigma_{14}\sigma_{34} - \omega^2 m_{14}m_{34})/rr_4 + pl \\ &+ pm_{13} - p(\sigma_{14}m_{34} + \sigma_{34}m_{14})/r_4] \\ &= r(1 - \epsilon + \alpha + \beta + \gamma + \delta) \quad . \quad (10) \end{aligned}$$

where ϵ is written for the term pl, which does not change during the measurements, and which can be precisely evaluated, while α, β, γ and δ, defined below, may change sign (and perhaps magnitude) as the various connections are changed.

$$\begin{aligned} \alpha &= \sigma_{13}/r \\ \beta &= -(\sigma_{14}\sigma_{34} - \omega^2 m_{14}m_{34})/rr_4 \\ \gamma &= pm_{13} \\ \delta &= -p(\sigma_{14}m_{34} + \sigma_{34}m_{14})/r_4 \end{aligned}$$

We have now to examine the behaviour of these factors as we work through the arrangements set out in Table 2. We can say quite definitely that all m terms will change sign as the appropriate connections are reversed: we cannot, however, readily decide what is going to happen to the σ terms, and for the moment we shall assign to them different values for the different connections. In Table 3 the possible variations are set out against the coil arrangements listed in Table 2: the composite terms β and δ have been written as $\beta' + \beta''$ and $\delta' + \delta''$ with appropriate suffixes.

Table 3

VALUES TO BE ASSIGNED TO α, β, γ, δ

Arrangement	α	β	γ	δ
1	α_1	$\beta'_1 + \beta''_1$	γ_1	$\delta'_1 + \delta''_1$
2	α_2	$\beta'_2 - \beta''_1$	$-\gamma_1$	$\delta'_2 - \delta''_1$
3	α_3	$\beta'_3 + \beta''_1$	γ_1	$-\delta'_2 - \delta''_2$
4	α_4	$\beta'_4 - \beta''_1$	$-\gamma_1$	$-\delta'_1 + \delta''_2$
5	α_4	$\beta'_4 - \beta''_1$	$-\gamma_1$	$-\delta'_1 + \delta_2$
6	α_3	$\beta'_3 + \beta''_1$	γ_1	$-\delta'_2 - \delta''_2$
7	α_2	$\beta'_2 - \beta''_1$	$-\gamma_1$	$\delta'_2 - \delta''_1$
8	α_1	$\beta'_1 + \beta''_1$	γ_1	$\delta'_1 + \delta''_1$
Means	$\frac{1}{4}\Sigma\alpha$	$\frac{1}{4}\Sigma\beta'$	0	0

Rayner (*loc. cit.*) made an empirical analysis of the systematic variations "on the assumption that the differences of each set from the mean were composed of a few terms, each being constant in value, but with a sign that depended on the sense in which certain coils were connected." The first part of this assumption is in accord with our analysis, but we take the matter further and suggest that these terms also vary in magnitude. In fact, it appears that there are four distinct effective values of r operating in the sequence of reversals, giving four pairs of results. Now the original observations of Hartshorn and Astbury do indeed agree closely with this—i.e. set 8 corresponds to set 1, set 7 to set 2, and so on. We cannot expect exact correspondences because of random errors of balance setting, and what, in

retrospect, can have been little better than random setting of the zero adjuster. Rayner's results[1] do not show the correspondence so markedly; the differences, given as parts in 10^6, between "corresponding" sets are summarized in Table 4.

Table 4

Sets	Hartshorn and Astbury	Rayner
1–8	$-1{\cdot}6$	11
2–7	$-8{\cdot}0$	3
3–6	$4{\cdot}4$	8
4–5	$-3{\cdot}8$	0

So far as the original results are concerned, if we assess the significance of the agreement by a statistical "t test," then, except for 2–7, which is bad, the agreement between the other groups is about the 5% level of significance, which we might accept as indicating that the group pairs are statistically identical. On the other hand, if we compare the overall means for sets 1 to 4 with those for sets 5 to 8, we get agreement to about the 25% level of significance. It seems very likely, then, that we are, as our theory predicts, really dealing not with eight separate results in each group, but with two sets of four, differentiated by random errors.

The mean of the eight sets involves the residual factors $\Sigma\alpha = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$ and $\Sigma\beta' = \beta_1' + \beta_2' + \beta_3' + \beta_4'$, which are all impurity terms, and to which, so far, we have ascribed no particular meaning. Now the α terms are variants of σ_{13}, which is the impurity of the mutual inductance between meshes 1 and 3, and, as defined, excludes the true eddy-current effect. It is therefore to be attributed to mutual capacitance between these circuits, and, using Butterworth's results,[8] it appears that the average value will be of the order of $\omega^2 C_{13} L' R'/r$, where C_{13} is the mutual capacitance of the circuits and L' and R' are the associated inductance and resistance. If we take as likely values $C_{13} = 25\,\mu\mu$F, $L' = 10$mH, $R' = 10$ ohms, this gives us a term of the order of 1 part in 10^6. Similarly, we may suppose the β' terms to depend upon capacitances between the coils and the conducting masses causing the eddy currents, and it is unlikely that these can contribute a factor greater than that we have just estimated. In calculating the actual ratio of the units, we take the square root of our equations, so that in the final answer any residuals of this kind are halved. We may say then that the mean of our two sets of four results is likely to be "true" to about 1 part in 10^6. There seems little point in trying to evaluate residuals from the published results: the most we can say is that the β and γ terms appear to contribute together about 3–5 parts in 10^6, and the δ terms about 6–8 parts in 10^6, but no great weight can be given to these estimates.

We have given what appears to be a valid theoretical explanation of the systematic variations observed on the Campbell bridge by Hartshorn and Astbury and by Rayner, and we have gone some way towards justifying the acceptance of the mean result of eight sets as "correct." It remains to offer a direct experimental demonstration of the effect, and this we shall do with reference to a "new" bridge circuit.

(3) A "NEW" BRIDGE

The Campbell circuit is essentially that of an a.c. potentiometer, and in order to establish satisfactory earthing conditions the inversion of the resistor r (Fig. 1) was proposed by Hartshorn. This still leaves the common point conditions of the mutual inductors a little obscure—indeed, Rayner points out that there is some uncertainty in the frequency variation of the inductances

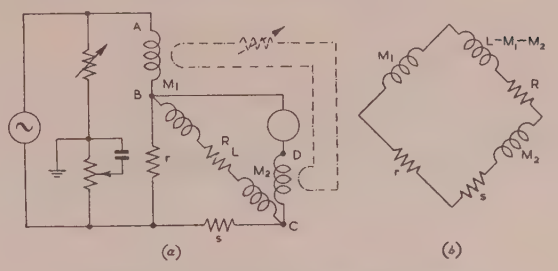


Fig. 3.—New bridge and transformation.

—and it is possible that a more compact bridge circuit may be worth investigation. The circuit of Fig. 3(*a*) is proposed. Ideally, this transforms into the circuit of Fig. 3(*b*), the balance conditions of which are

$$-\omega^2 M_1 M_2 = Rr$$
$$M_1 s = r(L - M_1 - M_2)$$

The mutual inductances must therefore be of opposite signs and M_1 must be positive (in the sense used by Hartshorn and Astbury): the working form of the ideal equations is thus

$$\omega^2 M_1 M_2 = Rr$$
$$M_1 s = r(L + M_2 - M_1)$$

It will be observed that two completely independent balance adjustments are provided by R and s, and not by M_2 and s as in the Campbell circuit. This is no drawback, however, for balance can be approached sufficiently closely on R and s and the final balance done on M_2 and s. Near to balance the convergence depends upon $\omega M_1/r$ being sufficiently greater than unity: for the Campbell bridge at 100c/s this value is about 6. A model of the new bridge has been made for operation at 1kc/s with $M_1 \simeq M_2 \simeq 1$mH, $r = 1$ ohm, so that the convergence factor would be the same as for a 100c/s bridge with $M_1 = M_2 = 10$mH and $r = 1$ ohm.

The general equations of the new bridge are most easily written down by transformations: it is useful to have available first the transformation shown in Fig. 4, which converts a 4-arm

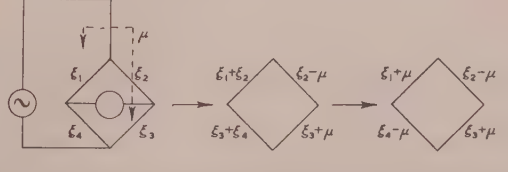


Fig. 4.—Bridge with coupling between supply and detector.

bridge with mutual impedance μ between supply and detector branches into a simple 4-arm net. Using this, our bridge transforms as shown in Fig. 5, giving as balance condition

$$(\mu_{12} + \mu_{13})(\xi_3 + \mu_{23} + \mu_{13}) = (\xi_1 - \mu_{13})(\xi_2 - \mu_{23} - \mu_{12} - \mu_{13})$$

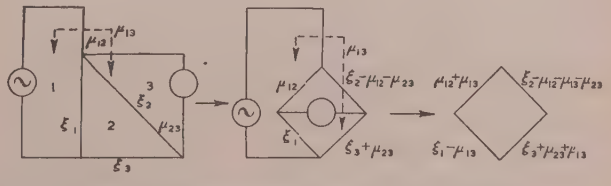


Fig. 5.—Complete transformation of new bridge.

which is readily recast as

$$(\mu_{12}+\mu_{13})(\xi_1+\xi_2+\xi_3)=(\xi_1+\mu_{12})(\xi_2-\mu_{23}) \quad . \quad (11)$$

Every term in this equation is to be interpreted as the effective value of the impedance represented in the presence of eddy currents: thus [cf. eqns. (1)–(4)] every μ contains a term of the type $\mu_{14}\mu_{24}/\xi_4$, and every ξ a term of the type μ_{24}^2/ξ_4, where the subscript 4 refers to a notional eddy-current circuit. Explicitly we write (remembering that M_1 and M_2 are of opposite sign)

$$\xi_1=r+j\omega l, \quad \xi_2=R+j\omega L, \quad \xi_3=s+j\omega\lambda$$

$$\mu_{12}=\sigma_1+j\omega M_1, \; -\mu_{23}=\sigma_2+j\omega M_2, \; \mu_{12}+\mu_{13}=\sigma_3+j\omega M_3$$

Substituting these values in eqn. (11), we find for the main equation of balance

$$\omega^2 M_1M_2\{1-[L(M_1-M_3)+l(M_2-M_3)+(lL+M_3\lambda)]/M_1M_2\}$$
$$=rR[1+\sigma_3 s/rR-(\sigma_1-\sigma_3)/r-(\sigma_2-\sigma_3)/R] \quad . \quad (12)$$

There are clearly four ways in which the coils of the inductors can be connected to secure balance: if eddy-current effects exist we shall get, correspondingly, four different settings of the bridge. But if, for every one of the four settings, we measure M_1 and σ_1 for the system ABC [Fig. 3(*a*)], M_2 and σ_2 for the system BCD and M_3 and σ_3 for ABD, and if our arguments are valid, there should be no systematic differences between the four results. We could indeed achieve the same effect with the Campbell circuit by determining, for every coil combination, the effective resistance and inductance of the resistor r (Fig. 1), using the main supply and detector terminals as current and potential terminals respectively, the connecting loop (NO, Fig. 1) being open-circuited. The present suggestion of measuring an additional mutual inductance seems less difficult to carry out, and the circuit requires little adaptation to permit the measurement of the three mutual inductances in turn against the primary standard or the substandard inductometer. It does indeed seem preferable to make the measurements directly in this way, rather than by measuring one M and then comparing the others with it.

(4) AN EXPERIMENTAL DEMONSTRATION OF THE EDDY-CURRENT EFFECT

It is beyond the author's present resources to attempt a precision investigation of the new bridge, but a direct demonstration of the eddy-current effect has been made on a model of the new bridge designed to operate at a frequency of 1 kc/s. The mutual inductors are toroidally-wound on wooden formers of mean diameter 10 cm and of square section of 5 cm side length. A cylindrical hole bored in one former accommodates a small circular coil which can be turned about an axis lying in the mean plane of the toroid and passing through its centre. This gives a variation in mutual inductance of about 7 μH. A "bridge centre" is built on a block of Perspex, 6 in × 3 in × $\frac{1}{4}$ in, and carries the resistors R, s and r, external connections being provided through which R and s can be shunted by ordinary decade boxes. The Perspex panel carries a pattern of blind holes, of $\frac{1}{8}$ in diameter and $\frac{3}{16}$ in deep, which are filled with mercury and are interconnected with links made from short lengths of No. 20 s.w.g. copper wire. The essentials of the arrangement are shown in Fig. 6. A Wagner arm and reversing switches for supply and detector are provided, the former (see Fig. 4) consisting of two radio-type potentiometers in series across the supply, their common point being earthed and a fixed capacitor being connected across the tapping of one of them. The absence of any massive metal parts and the toroidal structure of the inductors reduces to a very low level any adventitious eddy-current effects, and "artificial" eddy currents are provided and controlled by tertiary windings on each inductor, which are connected in series through an adjustable resistor, as indicated by the chain-dotted circuit in

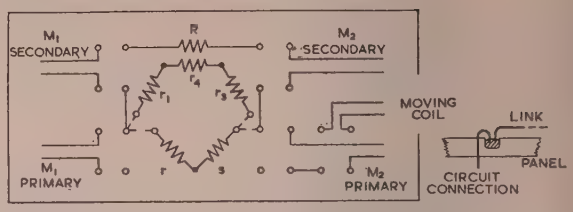


Fig. 6.—Bridge centre: supply and detector details omitted.

Fig. 3(*a*). The main windings of the inductors consist of about 300 uniformly-spaced turns: the tertiary windings have 10 turns, closely packed.

The bridge is fed from a push-pull amplifier driven by a resistance-capacitance audio signal-generator; bridge currents up to 350 mA were used. The detector is a 3-stage amplifier with twin-T feedback, giving adequate discrimination against harmonics, and balance is observed with an ordinary telephone, a cathode-ray oscillograph serving as a monitor. The sensitivity was such that a change of about 5 parts in 10^5 could be detected on M_2. The bridge was initially balanced as closely as possible by shunting R and s, the final balance being taken up on the moving coil of M_2. Reversals of supply and detector connections produced negligible effects, and the simple system of links made it easy to secure all possible coil connections in quick succession without introducing changes of resistance or of stray inductance. There was sometimes a slight drift in frequency, but this was never serious. In taking a series of readings, the tertiary circuit was first opened and the bridge balance secured for the four possible coil connections in rapid succession. In this condition, there are, effectively, no eddy-current circuits, and the four balance points were always identical down to the limits of sensitivity. The tertiary circuit was then closed through a known resistance and the four balance points again observed. Here the balance points fell into two nearly identical pairs, the "spread" increasing with the conductance of the tertiary circuit. For every setting of resistance in the tertiary circuit the effective value of μ_{13} was measured on an auxiliary bridge, built on to the main bridge centre and brought into operation by simple link and switch changes. This bridge is shown in Fig. 7. The

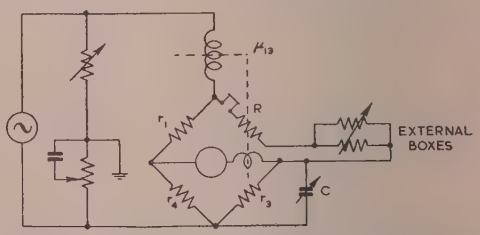


Fig. 7.—Auxiliary bridge for μ_{13}: tertiary circuit coupling indicated by dotted line.

secondary winding of M_1 is replaced by a link and that of M_2 by external resistance boxes: wire-wound resistors r_1, r_3 and r_4 (Figs. 6 and 7), each of 100 ohms, complete the bridge, one of the arms being shunted by a variable air-capacitor, C. A Wagner arm, similar to that in the main bridge, is provided. Applying the transformation of Fig. 4 to the bridge of Fig. 7, we have, at balance, if $\mu_{13}=\sigma+j\omega m$, and r_2 is the total resistance of the R arm,

$$\sigma=\frac{r_1r_3-r_2r_4}{r_1+r_2+r_3+r_4}$$

$$\omega m=\frac{(\alpha_1+\alpha_3-\alpha_2-\alpha_4)r_1r_3}{r_1+r_2+r_3+r_4}$$

where all the α's are the phase angles of the arms indicated by the subscripts. The zero of μ_{13} can be taken either with the secondary winding of M_2 replaced by a link or with the tertiary circuit open. We then have for any other setting

$$\sigma_{eff} = \frac{r_4 \Delta r_2}{r_1 + r_2 + r_3 + r_4} \simeq \tfrac{1}{4}\Delta r_2$$

$$m_{eff} = \frac{r_1 r_3 \Delta\alpha_3}{\omega(r_1 + r_2 + r_3 + r_4)} \simeq \tfrac{1}{4} r_3^2 \, \Delta C$$

where σ_{eff} and m_{eff} are the actual values of the components of μ_{13} with which we are concerned and Δr_2 and ΔC are respectively the changes made in r_2 and C between the zero condition for μ_{13} and any other setting. As might have been expected from the experimental arrangements, m_{eff} remained constant at about 0·01 μH, corresponding to the stray mutual inductance between the primary windings of M_1 and M_2; by introducing into the tertiary circuit an inductance of about 1 mH (wound on a miniature Ferroxcube core) m_{eff} can be changed. However, the experiments showed that m_{eff} alone had little effect on the changes arising from the coil reversals, the situation being dominated almost entirely by σ_{eff}. The results of a number of trials are summarized in Figs. 8 and 9. In Fig. 8 the moving-coil reading

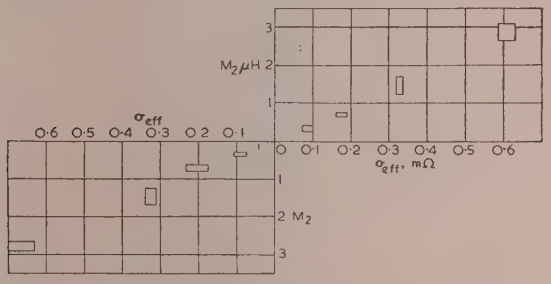


Fig. 8.—Correlation of M_2 readings at balance with σ_{eff}.

Upper-quadrant coil connections

Primary		Secondary	
M_1	M_2	M_1	M_2
+	+	+	+
−	−	−	−

Lower-quadrant coil connections

Primary		Secondary	
M_1	M_2	M_1	M_2
−	+	−	+
+	−	+	−

of M_2 is plotted as a function of the measured value of σ_{eff}, the observed points being scattered within the small rectangles. In Fig. 9 the mean readings of M_2 and of σ_{eff} are plotted against the conductance of the tertiary circuit: this shows very clearly how the spread of M_2 (and hence, on our bridge, the spread in the ratio of the absolute to the international ohm) follows the incidence of uncompensated eddy-current effects, represented by σ_{eff}.

(5) CONCLUSIONS

It has been shown analytically and demonstrated experimentally on an equivalent circuit that the systematic errors

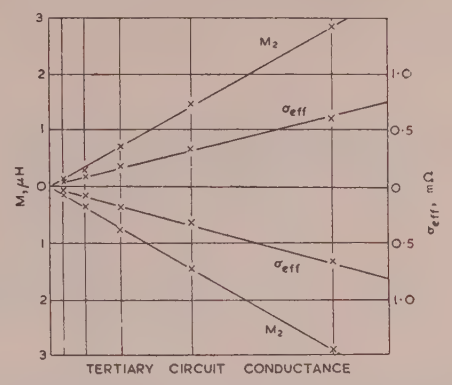


Fig. 9.—Variation of M_2 and σ_{eff} with tertiary circuit con ductance.

associated with coil reversals in the Campbell a.c. bridge for the absolute measurement of resistance can be explained by uncompensated eddy-current coupling between supply and detector meshes. It is also suggested that the zero-adjusting system in the Hartshorn–Astbury version of the Campbell bridge was largely ineffective. The new bridge proposed in the present paper can be operated to eliminate the eddy-current effect without external compensation if three mutual impedances (two of which are nearly equal) are determined instead of two, as in the Campbell bridge. If this bridge should be developed for precision work, it is suggested that the link system used for the coil connections could very well be on a less massive scale than that used by Hartshorn and Astbury.

It may well be argued that if the eddy-current hypothesis is correct, the systematic error effect should be less at a frequency of 50 c/s than at a frequency of 100 c/s. The results of Hartshorn and Astbury offer no real evidence on this point: a comparison can be made only if it is known that the eddy-current conditions (and in particular the settings of the zero adjuster) were the same for the two frequencies.

(6) ACKNOWLEDGMENTS

It is a pleasure to record appreciation of the assistance of Mr. P. F. Ridler, B.E., and Mr. D. P. Saville, M.Sc., in providing the supply and detector amplifiers.

(7) REFERENCES

(1) Rayner, G. H.: "An Absolute Measurement of Resistance by Albert Campbell's Bridge Method," *Proceedings I.E.E.*, Monograph No. 95 M, April, 1954 (**101**, Part IV, p. 251).

(2) Vigoureux, P.: *Collected Research Reports of the National Physical Laboratory*, London, **24**, Paper No. 8.

(3) Grüneisen and Giebe: *Annalen der Physik* (*Leipzig*), 1920, **368**, p. 179.

(4) Curtis, H. L., Moon, C., and Sparks, C. M.: *Journal of Research of the National Bureau of Standards*, 1936, **16**, p. 1.

(5) Thomas, J. L., Peterson, C., Cooter, I. L., and Kotter, F. R.: *ibid.*, 1949, **43**, p. 291.

(6) Hartshorn, L., and Astbury, N. F.: *Philosophical Transactions of the Royal Society*, A, 1937, **236**, p. 423.

(7) Jouaust, R., Picard, M., and Hérou, R.: *Bulletin de la Société Française des Electriciens*, 1938.

(8) Butterworth, S.: *Proceedings of the Physical Society*, 1921, **33**, p. 312.

621.311.44:621.3.015.1:621.3.018.14

Monograph No. 131 S
May 1955

ALTERNATING VOLTAGE, DIRECT-VOLTAGE REGULATION AND POWER FACTOR OF CONVERTOR STATIONS OPERATING ON A.C. SYSTEMS OF FINITE SHORT-CIRCUIT CAPACITY

By ERICH UHLMANN, Dr.Ing.

(*The paper was first received* 24*th June*, 1954, *and in revised form* 25*th January*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in May*, 1955.)

SUMMARY

The paper deals with the effect of the impedance of the a.c. system on the voltage and power factor at the a.c. terminals of a convertor such as a mercury-arc-rectifier group, and on the direct-voltage regulation of this group. For convenience, the results are given as charts of curves, which are also being incorporated in a forthcoming I.E.C. publication, and the paper gives the derivation of these. For the investigation an equivalent connection diagram is used, the impedance of the a.c. system, including that of the generators feeding the system, being shown on this. A formula is derived relating the r.m.s. value of the alternating voltage, U_L, in the convertor station and that at the imaginary feeding point, where the voltage, even with convertor loading, remains sinusoidal. The results obtained are then used in the deduction of the additional direct-voltage regulation, due to the a.c. system, which must be introduced on the d.c. side if the conventional theory of convertors has been used to calculate the direct voltage, thereby the convertor-station voltage being assumed to be sinusoidal and of r.m.s. value equal to U_L. Curves are given showing the voltage regulation over a range covering most practical cases. A formula for calculating the fundamental wave of the alternating voltage in the convertor station is also derived in the paper. By means of this formula, together with that for the r.m.s. value of the convertor-station alternating voltage, it is shown that the displacement factor and power factor in the convertor station are determined by factors which depend entirely on known plant and a.c.-system data, and consequently that acceptance-test measurements of power factor in convertor stations actually serve no useful purpose, since they merely confirm this dependency.

LIST OF SYMBOLS*

I_d = Arithmetic mean value of the direct current.
I_L = R.M.S. value of the line current on the a.c. side of the convertor station.
I_{Li} = R.M.S. value of the ideal line current on the a.c. side of the convertor station, computed from the direct current on the basis of rectangular anode-current waveshapes.
P = Active power at the convertor station, on the a.c. side.
Q = Three-phase short-circuit capacity of the a.c. system.
S = Apparent power at the convertor station, on the a.c. side.
S_F = Apparent power of the fundamental wave of the alternating current in the convertor station.
U_L = R.M.S. value of the line voltage on the a.c. side of the convertor station.
U_{di0} = Arithmetic mean value of the ideal no-load direct voltage, computed on the basis of a sinusoidal line voltage on the a.c. side having an r.m.s. value equal to U_L.
d_t = Direct-voltage regulation due to the convertor main transformer and interphase transformer (if any), per unit of $U_{di0}\,I_d$.
d_b = Direct-voltage regulation due to other parts of the convertor, e.g. reactors, etc., per unit of $U_{di0}I_d$.
d_L = Additional direct-voltage regulation due to the impedance of the a.c. system, per unit of $U_{di0}I_d$.
e_{xL}, e_{rL} = Reactive and resistive short-circuit voltages, respectively, of the a.c. system, per unit of $U_{di0}I_d$.
$\cot \phi_L$ = Ratio of resistance to reactance of the a.c. system.
$\cos \phi$ = Displacement factor, i.e. ratio of active power to apparent power of the fundamental wave of the alternating current, at the convertor station.
λ = Power factor, i.e. ratio of active power to apparent power on the a.c. side of the convertor station.
α = Delay angle.
α_s = Spontaneous delay angle.
ψ = Phase angle.
θ = Electrical time angle.
u = Angle of overlap.
p = Pulse number.

Suffices

0 = At no load.
1 = Values at rated direct current, I_{d1},
b = Reactors.
d = Direct current or voltage.
i = Ideal.
L = Line (a.c. system).
r = Resistive.
t = Transformer.
x = Reactive.
y = Any numeral.
F = Fundamental-wave value.

A prime appended to a symbol denotes that the value refers to conditions at the imaginary feeding point.

The lower-case symbols i_L, u_L, etc., are used to indicate instantaneous currents and voltages on the a.c. side.

Various other symbols, e.g. ρ, χ, etc., are introduced throughout the text to replace more complicated mathematical expressions; the definitions of these symbols are in all cases given implicitly by the substitutions in question.

(1) INTRODUCTION

A new document, "Recommendations for Mercury-Arc Convertors," which is expected to be published shortly by the International Electrotechnical Commission, contains a number of novel charts which give the effect of the a.c.-system impedance on the direct-voltage regulation and power factor of a mercury-arc convertor station. The purpose of the paper is to show the derivation of the formulae by which these curves have been calculated and to extend the treatment of these questions, and at

* The symbols used in the paper agree with those in the forthcoming I.E.C. publication, "Recommendations for Mercury-Arc Convertors."

Correspondence on Monographs is invited for consideration with a view to publication.
Dr. Uhlmann is with Allmänna Svenska Elektriska Aktiebolaget, Sweden.

the same time to draw attention to some conclusions that follow regarding the power factor of such convertors.

In order to study the extent to which the impedance of the a.c. system affects the alternating voltage and the power factor in a convertor station, an equivalent connection diagram, as shown in Fig. 1, has been assumed. The generator shown in this Figure

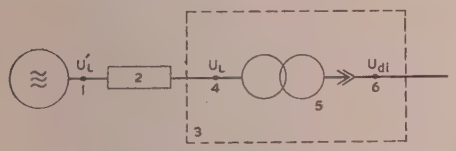


Fig. 1.—Equivalent connection diagram for a convertor plant.

1. Imaginary feeding point.
2. Impedance of a.c. network, including that of the generators.
3. Convertor station.
4. A.C. terminals of the convertor station.
5. Convertor transformer and valves.
6. D.C. terminals of the convertor station.

is an idealized one, in that it delivers at the "feeding point," 1, a sinusoidal voltage, with r.m.s. value U'_L, independent of the nature of the load current drawn from it. The equivalent impedance of the a.c. system between the feeding point and the a.c. terminals, 4, of the convertor station, 3, is represented by a lumped impedance, 2, in the Figure. This impedance also contains a fraction due to the internal impedances of the generators feeding the system. The r.m.s. value of the voltage at the a.c. terminals of the convertor station is denoted by U_L; this voltage is, in general, non-sinusoidal, but it becomes sinusoidal in the special case where the convertor loading is zero, when $U_L = U'_L$. In the convertor station are placed the convertor transformer and the valves, 5. The voltage at the d.c. terminals, 6, of the convertor station is denoted by U_{dl}. As usual, it is assumed that the ripple in the direct current of the convertor is negligibly small, and that the arc drop has a constant value (which depends on the direct current).

The conventional theory of convertors gives formulae relating the voltage at the d.c. terminals with that at the feeding point, but it does not give any formulae relating the former voltage with that at the a.c. terminals of the convertor station itself. It is obvious that the voltage at the imaginary feeding point cannot, in general, be measured, whereas the r.m.s. value of the voltage at the a.c. terminals can quite easily be measured, being, in fact, the rated alternating voltage at that point. This latter voltage will also be the value measured when acceptance tests are being carried out. In the present paper a formula is derived which gives the ratio of the alternating voltage at the convertor station to that at the imaginary feeding point, this ratio, U_L/U'_L, being given as a function of known plant and a.c.-system data. In the calculation of this ratio it is assumed that the voltage at the feeding point is varied in such a way as to maintain the r.m.s. value of the alternating voltage at the convertor station constant. The actual value of this ratio for most practical cases is shown graphically.

In Section 3 the practical calculation of convertor plant based on the rated voltage at the a.c. terminals of the station has been simplified appreciably. One merely has to calculate the direct voltage, using the conventional formulae with the assumption that U_L is sinusoidal; an additional drop, d_L, has then to be subtracted from this in order to account for the impedance of the a.c. system. The formula for this additional voltage drop is derived in the paper, and its value for most practical cases is shown graphically.

To evaluate the displacement factor at the convertor station itself, it is necessary to know the value of the fundamental wave of the alternating voltage, U_{LF}, at this point. A formula for this is derived in the paper, but since its value differs very little from the r.m.s. value of the alternating voltage at the same point, no graphs showing this voltage are given.

The final Section of the paper deals with the power factors in convertor stations. The conventional theory of convertor power-factor is valid only at the imaginary feeding point, whereas in other electrical apparatus the power factor is understood as being the value which exists at the terminals of the apparatus itself. Since only the latter power factor can be measured during acceptance tests, it seems more suitable that the general convention should also be adopted for convertor stations.

However, the power factors of convertors seem to have been given an unjustifiable amount of attention. If the magnetizing current is neglected in both cases, the power factor of a convertor—like that of a transformer—is determined by data which are fixed by other, definite, factors. For a transformer these are the short-circuit voltage and the power factor of the load; for a convertor they are the direct-voltage regulation and the control angle. It can therefore be assumed that, in the future, acceptance-test measurements of convertor power-factors will be omitted, and that it will be taken for granted that these have values which are determined entirely by known plant and a.c.-system data—as is the accepted practice for transformers.

In the meantime, until such an opinion has been more generally accepted, it may be of interest to know the power factors in convertor stations. Formulae are therefore derived for the power factor, λ, and the displacement factor, $\cos\phi$, in the convertor station itself, and in them the effect of the impedance of the a.c. system has been taken into account. Graphs are given showing the values of these over a range covering most practical cases. Approximate formulae are also indicated, from which these same quantities can be determined to an accuracy of $\pm 0{\cdot}01$.

(2) THE R.M.S. VALUE OF THE ALTERNATING VOLTAGE AT THE CONVERTOR STATION*

The instantaneous value of the alternating voltage at the convertor station is obtained from the equation:

$$\frac{u_L}{U'_L} = \frac{u'_L}{U'_L} - e'_{rL}\frac{i_L}{I_{Li}} - e'_{xL}\frac{di_L/d\theta}{I_{Li}} \quad . \quad . \quad . \quad (1)$$

The r.m.s. value of this voltage then follows from the expression

$$\frac{U_L}{U'_L} = \left[\frac{1}{2\pi}\int_0^{2\pi}\left(\frac{u_L}{U'_L}\right)^2 d\theta\right]^{1/2} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The substitution of eqn. (1) in eqn. (2) gives six integrals, (*a*)–(*f*), the values of which are as follow:

(*a*) $$\frac{1}{2\pi}\int_0^{2\pi}\left(\frac{u'_L}{U'_L}\right)^2 d\theta = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

(*b*) $$\frac{1}{2\pi}\int_0^{2\pi}\left(\frac{i_L}{I_{Li}}\right)^2 d\theta = 1 - c\psi(\alpha, u) \quad . \quad . \quad . \quad (4)$$

where $$c = 2p\sin^2\frac{\pi}{p} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

i.e. $$c_{p=6} = 3 \text{ and } c_{p=12} = 1{\cdot}6077$$

* Uhlmann, E., and Forssell, H.: "Beräkning av effektfaktorerna och transformatorns primära belastningsspänning i de svenska strömriktarnormerna," *Teknisk Tidskrift*, 1941, 71, No. 9, p. 41.

and

$$\psi(\alpha, u) = \frac{1}{8\pi(d_x')^2}\{[2 + \cos(2\alpha + u)]\sin u - u[1 + 2\cos\alpha\cos(\alpha + u)]\} \quad . \quad (6)$$

where $$d_x' = \tfrac{1}{2}[(\cos\alpha - \cos(\alpha + u)] \quad . \quad . \quad . \quad . \quad (7)$$

(c) $$\frac{1}{2\pi}\int_0^{2\pi} \frac{u_L' i_L}{U_L' I_{Li}} d\theta = \left(\frac{p}{\pi}\sin\frac{\pi}{p}\right)\rho(\alpha, u) \quad . \quad . \quad . \quad (8)$$

where $$\rho(\alpha, u) = \tfrac{1}{2}[\cos\alpha + \cos(\alpha + u)] \quad . \quad . \quad . \quad (9)$$

(d) $$\frac{1}{2\pi}\int_0^{2\pi} \frac{u_L'}{U_L' I_{Li}}\frac{di_L}{d\theta} d\theta = \left(\frac{p}{\pi}\sin\frac{\pi}{p}\right)\chi(\alpha, u) \quad . \quad . \quad (10)$$

where $$\chi(\alpha, u) = \frac{1}{8d_x'}[2u + \sin 2\alpha - \sin 2(\alpha + u)] \quad . \quad (11)$$

Integrals (*b*), (*c*) and (*d*) follow from the well-known relationships given by the conventional theory of convertors regarding the r.m.s. value of the alternating current and the active and reactive powers at the feeding point.

The fifth integral is

(e) $$\frac{1}{2\pi}\int_0^{2\pi} \frac{i_L}{U_L' I_{Li}}\frac{di_L}{d\theta} d\theta = 0 \quad . \quad . \quad . \quad . \quad . \quad (12)$$

In order to evaluate the final integral, use is made of the fact that the alternating current on the a.c. line side consists, during the intervals between commutations, of steps which lie on a sine wave. Hence this current can be denoted by

$$i_{Ly} = \sqrt{2} I_{Li}\cos\psi_y \quad . \quad . \quad . \quad . \quad . \quad (13)$$

where $$\psi_{y+1} - \psi_y = \frac{2\pi}{p} \quad . \quad . \quad . \quad . \quad . \quad (14)$$

During the transition from the current step y to the step $y + 1$, the line current of the a.c. system can be written as

$$i_L = \sqrt{2} I_{Li}\{[1 - w(\theta)]\cos\psi_y + w(\theta)\cos\psi_{y+1}\} \quad . \quad (15)$$

where $$w(\theta) = \frac{\cos\alpha - \cos(\theta - \psi_y)}{\cos\alpha - \cos(\alpha + u)} \quad . \quad . \quad . \quad (16)$$

is the well-known transition function, which at the beginning ($\theta = \psi_y + \alpha$) and end ($\theta = \psi_y + \alpha + u$) of the commutation has the values 0 and 1 respectively.

The final integral thus becomes:

(f) $$\frac{1}{2\pi}\int_0^{2\pi/p} \frac{\left(\frac{di_L}{d\theta}\right)^2}{I_{Li}^2} d\theta = \frac{1}{2\pi}\sum_{y=1}^{p}\int_{\psi_y+\alpha}^{\psi_y+\alpha+u} \frac{\left(\frac{di_L}{d\theta}\right)^2}{I_{Li}^2} d\theta$$

$$= \frac{1}{2\pi}\sum_{y=1}^{p} 8\sin^2\frac{\pi}{p}\sin^2\left(\psi_y + \frac{\pi}{p}\right)\int_{\psi_y+\alpha}^{\psi_y+\alpha+u} \frac{\sin^2(\theta - \psi_y)}{[\cos\alpha - \cos(\alpha + u)]^2} d\theta$$

$$= \frac{p}{\pi}\sin^2\frac{\pi}{p}\frac{\chi(\alpha, u)}{d_x'} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

If the above values of these integrals are now substituted in eqn. (2), the following expression is obtained:

$$\frac{U_L}{U_L'} = \left\{1 - \frac{p}{\pi}d_{xL}\frac{U_L}{U_L'}\left[\left(2 - \frac{d_{xL}}{d_x}\right)\chi(\alpha, u) + 2\cot\phi_L\rho(\alpha, u)\right] + \left(\frac{d_{xL}}{\sin\frac{\pi}{p}}\cot\phi_L\frac{U_L}{U_L'}\right)^2[1 - c\psi(\alpha, u)]\right\}^{\frac{1}{2}} \quad . \quad (18)$$

where $$d_{xL} = e_{xL}\sin\frac{\pi}{p} \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

$$d_x = d_{xt} + d_{xb} + d_{xL} \quad . \quad . \quad . \quad . \quad . \quad (20)$$

and $$\cot\phi_L = \frac{e_{rL}}{e_{xL}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

Eqn. (18) is of the second degree and from it the value of U_L/U_L' can easily be solved. An approximate solution can be obtained from the relationship

$$\frac{U_L}{U_L'} \simeq \frac{1}{1 + \frac{p}{2\pi}d_{xL}\frac{U_L}{U_L'}\left[\left(2 - \frac{d_{xL}}{d_x}\right)\chi(\alpha, u) + 2\cot\phi_L\rho(\alpha, u)\right]} \quad . \quad . \quad . \quad . \quad (22)$$

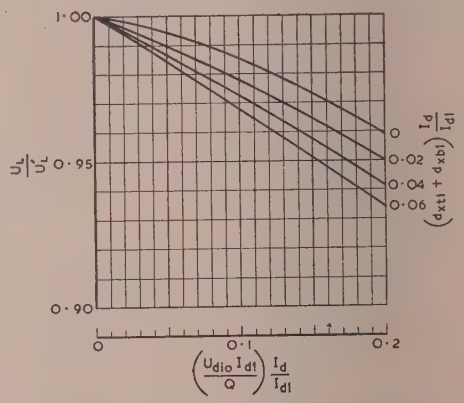


Fig. 2.—Ratio of r.m.s. value of alternating voltage at convertor station to that at the feeding point, for $p = 6$ and $\cot\phi_L = 0$.

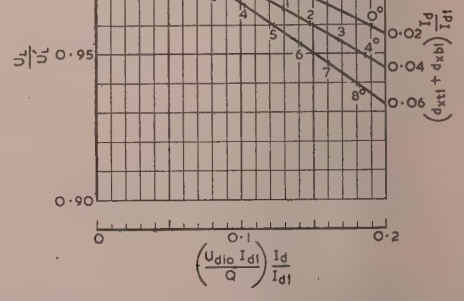


Fig. 3.—Ratio of r.m.s. value of alternating voltage at convertor station to that at the feeding point, for $p = 12$ and $\cot\phi_L = 0$.

α_s = Spontaneous delay angle.

A value obtained from expression (22) may then be substituted in the right-hand side of eqn. (18) in order to obtain a more accurate value, but the correction is negligible in most practical cases. Figs. 2 and 3 show values of U_L/U'_L calculated for $p = 6$ and $p = 12$. These curves refer to an a.c. system without resistance ($\cot \phi_L = 0$) and a rectifier without phase control ($\alpha = 0$).

In 12-pulse rectification the linear portion of the direct-voltage characteristic is departed from within certain ranges. As is generally known, a step in the valve voltage due to a commutation in one 6-pulse group causes a firing delay in the other group. This spontaneous firing delay of a rectifier without external phase control is denoted analogously as the spontaneous delay angle, α_s, and its value is given on the relevant curves in Fig. 3.

(3) THE ADDITIONAL DIRECT-VOLTAGE DROP DUE TO THE IMPEDANCE OF THE A.C. SYSTEM

If the arc drop in the valves is neglected, the direct voltage is obtained from the expression

$$U_{di} = U_{di0}\frac{U'_L}{U_L}(\cos\alpha - d'_t - d'_b - d'_{xL} - d'_{rL})$$

$$= U_{di0}\left[\frac{U'_L}{U_L}\cos\alpha - d_t - d_b - d_{xL}\left(1 + \frac{\cot\phi_L}{\frac{p}{\pi}\sin^2\frac{\pi}{p}}\right)\right] \quad . \quad (23)$$

where use has been made of the relationship

$$d_{rL} = e_{rL}\frac{\sqrt{(3)}U_L I_{Li}}{U_{di0} I_d} = \frac{e_{rL}}{\frac{p}{\pi}\sin\frac{\pi}{p}} \quad . \quad . \quad . \quad (24)$$

The additional voltage drop, d_L, due to the impedance of the a.c. system can now be obtained directly:

$$d_L = \frac{U_{di0} - U_{di}}{U_{di0}} - d_t - d_b$$

$$= 1 - \frac{U'_L}{U_L}\cos\alpha + d_{xL}\left(1 + \frac{\cot\phi_L}{\frac{p}{\pi}\sin^2\frac{\pi}{p}}\right) \quad . \quad (25)$$

For rectifiers without phase control or with spontaneous firing delay, this voltage drop is due entirely to the impedance of the a.c. system, and thus represents the additional direct-voltage drop due to this factor. For externally phase-controlled rectifiers, eqn. (25) can be rewritten in such a way that the effect of the phase control can be obtained separately, but this is of no interest in the present connection. Fig. 4 shows the additional voltage drop for $p = 6$ and $p = 12$. In order to show that the resistance of the a.c. system produces only a minor influence, the dotted curves are given for $\cot\phi_L = 0{\cdot}2$. As can be seen, the increase in voltage regulation due to the resistance factor is almost negligible, especially in 12-pulse operation. For this reason only one dotted curve has been shown for $p = 12$, namely that for $(d_{xt1} + d_{xb1})I_d/I_{d1} = 0$. In certain cases the additional voltage regulation can assume negative values, which means that the impedance of the a.c. system has a greater influence on the r.m.s. value of the station alternating voltage, U_L, than it has on the mean value of the direct voltage, U_{di}.

(4) THE FUNDAMENTAL WAVE OF THE ALTERNATING VOLTAGE AT THE CONVERTOR STATION

At the imaginary feeding point the fundamental wave of the alternating current has an active component

$$\frac{I_{LFr}}{I_{Li}} = \frac{p}{\pi}\sin\frac{\pi}{p}\rho(\alpha, u) \quad . \quad . \quad . \quad . \quad (26)$$

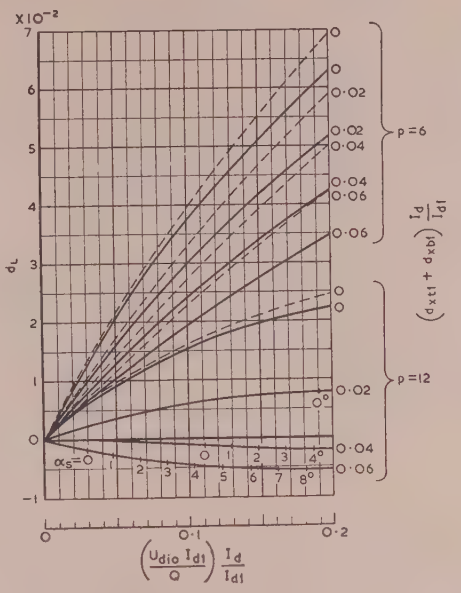


Fig. 4.—Additional direct voltage regulation drop due to a.c.-system impedance, for $p = 6$ and $p = 12$.

———— $\cot\phi_L = 0$.
– – – – – – $\cot\phi_L = 0{\cdot}2$.
α_s = Spontaneous delay angle.

and a reactive component

$$\frac{I_{LFx}}{I_{Li}} = \frac{p}{\pi}\sin\frac{\pi}{p}\chi(\alpha, u) \quad . \quad . \quad . \quad . \quad (27)$$

The fundamental wave of the alternating voltage in the convertor station is then obtained from the equation

$$\frac{U_{LF}}{U_L} = \left[\left(\frac{U'_L}{U_L} - e_{xL}\frac{I_{LFx}}{I_{Li}} - e_{rL}\frac{I_{LFr}}{I_{Li}}\right)^2 + \left(e_{xL}\frac{I_{LFr}}{I_{Li}} - e_{rL}\frac{I_{LFx}}{I_{Li}}\right)^2\right]^{1/2}$$

$$= \left\{\left(\frac{U'_L}{U_L}\right)^2 - 2\frac{p}{\pi}d_{xL}\frac{U'_L}{U_L}[\chi(\alpha, u) - \cot\phi_L\,\rho(\alpha, u)]\right.$$

$$\left. + \left(\frac{p}{\pi}\frac{d_{xL}}{\sin\phi_L}\right)^2[\rho^2(\alpha, u) + \chi^2(\alpha, u)]\right\}^{1/2} \quad . \quad . \quad . \quad . \quad (28)$$

The value of this voltage, U_{LF}, differs very little from the r.m.s. value, U_L, calculated in Section 2, the difference being especially small in 12-pulse rectification. On this account no special graphs showing this voltage have been given, since Figs. 2 and 3 give the desired values to a sufficient degree of accuracy.

(5) POWER FACTOR AND DISPLACEMENT FACTOR AT THE CONVERTOR STATION

If the magnetizing current of the transformer is neglected, the apparent power, S, at the convertor station is given by

$$S = \sqrt{(3)}U_L I_{Li}\sqrt{[1 - c\psi(\alpha, u)]} \quad . \quad . \quad . \quad (29)$$

and the active power, P, by

$$P = U_{di0} I_d\left[\frac{U'_L}{U_L}\rho(\alpha, u) - d_{rL}\right] \quad . \quad . \quad . \quad (30)$$

The apparent fundamental-wave power, S_F, follows from the relationship

$$S_F = \sqrt{(3)}U_{LF} I_{Li}\frac{p}{\pi}\sin\frac{\pi}{p}\sqrt{[\rho^2(\alpha, u) + \chi^2(\alpha, u)]} \quad . \quad (31)$$

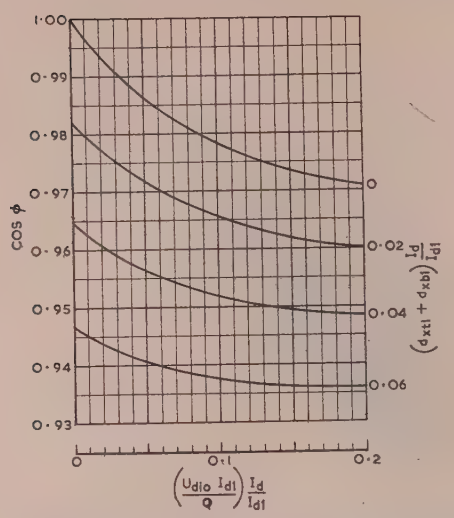


Fig. 5.—Displacement factor of convertor, neglecting magnetizing current of transformer, for $p = 6$ and $\cot \phi_L = 0$.

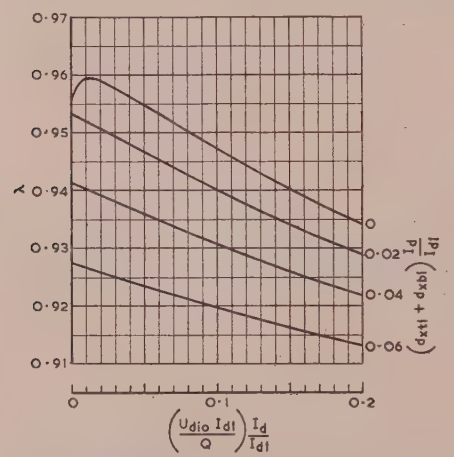


Fig. 6.—Power factor of convertor, neglecting magnetizing current of transformer, for $p = 6$ and $\cot \phi_L = 0$.

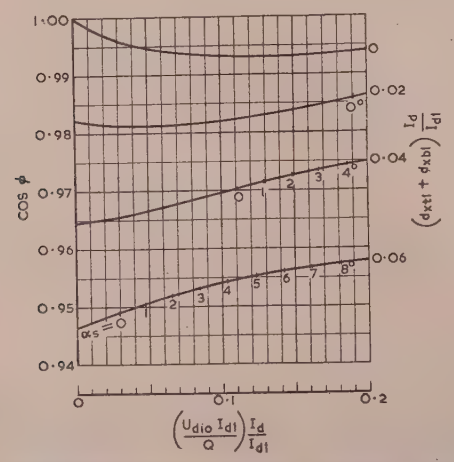


Fig. 7.—Displacement factor of convertor, neglecting magnetizing current of transformer, for $p = 12$ and $\cot \phi_L = 0$.

α_s = Spontaneous delay angle.

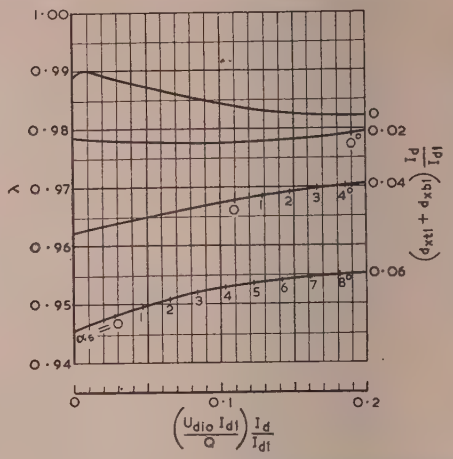


Fig. 8.—Power factor of convertor, neglecting magnetizing current of transformer, for $p = 12$ and $\cot \phi_L = 0$.

α_s = Spontaneous delay angle.

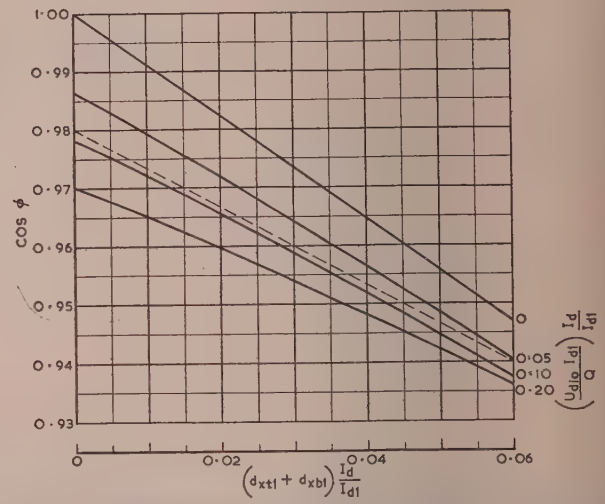


Fig. 9.—Displacement factor of convertor, neglecting magnetizing current of transformer, for $p = 6$ and $\cot \phi_L = 0$.

Approximate formula shown dotted.

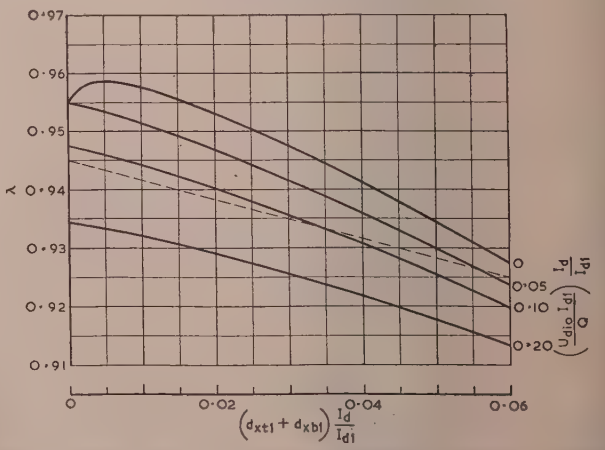


Fig. 10.—Power factor of convertor, neglecting magnetizing current of transformer, for $p = 6$ and $\cot \phi_L = 0$.

Approximate formula shown dotted.

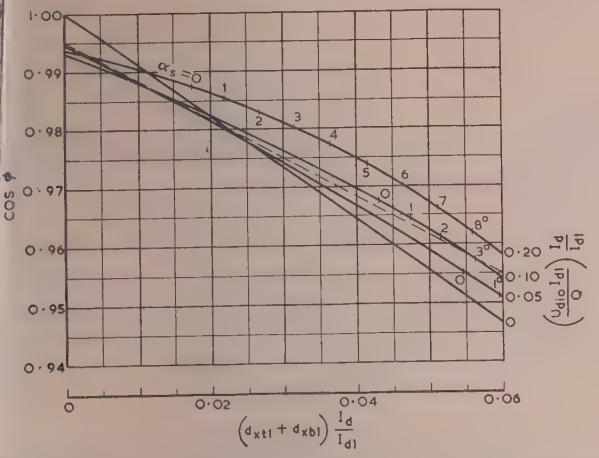
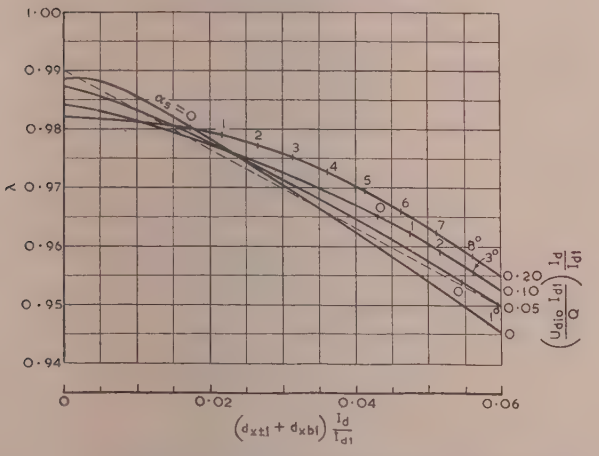


Fig. 11.—Displacement factor of convertor, neglecting magnetizing current of transformer, for $p = 12$ and $\cot \phi_L = 0$.
Approximate formula shown dotted.
α_s = Spontaneous delay angle.



ig. 12.—Power factor of convertor, neglecting magnetizing current of transformer, for $p = 12$ and $\cot \phi_L = 0$.
Approximate formula shown dotted.
α_s = Spontaneous delay angle.

The power factor, λ, at the convertor station is then

$$\lambda = \frac{P}{S} = \frac{p}{\pi} \sin \frac{\pi}{p} \frac{\rho(\alpha, u) - d_{rL}(U_L/U_L')}{\sqrt{[1 - c\psi(\alpha, u)]}} \frac{U_L'}{U_L} \quad . \quad (32)$$

nd the displacement factor, $\cos \phi$, is

$$\cos \phi = \frac{P}{S_F} = \frac{\rho(\alpha, u) - d_{rL}(U_L/U_L')}{\sqrt{[\rho^2(\alpha, u) + \chi^2(\alpha, u)]}} \frac{U_L'}{U_{LF}} \quad . \quad . \quad (33)$$

It follows from eqns. (32) and (33) that, if the resistive losses in the a.c. system are neglected, the power factor and the displacement factor at the convertor station are greater than those at the feeding point in the same proportion as the r.m.s. and fundamental values, respectively, of the alternating voltage at the feeding point are greater than those voltages at the convertor station. The values of λ and $\cos \phi$ are shown in Figs. 5–8 as a function of $(U_{di0}I_{d1}/Q)I_d/I_{d1}$ for $p = 6$ and $p = 12$. For these curves it has been assumed that the resistance of the a.c. system is negligible, i.e. $\cot \phi_L = 0$. In Figs. 9–12 the same results have been rearranged to show λ and $\cos \phi$ as a function of $(d_{xt1} + d_{xb1})I_d/I_{d1}$, with $(U_{di0}I_{d1}/Q)I_d/I_{d1}$ as parameter. These latter curves show that the power factors are only very slightly dependent on the a.c. system data, especially in 12-pulse operation. From these curves it follows that the power factor and displacement factor in convertor stations can, in most practical cases, be calculated with a discrepancy of less than $\pm 0{\cdot}01$ from the following four simple approximate formulae:

For $p = 6$:

$$\cos \phi \simeq 0{\cdot}980 - \tfrac{2}{3}(d_{xt1} + d_{xb1})\frac{I_d}{I_{d1}} \quad . \quad . \quad (34)$$

$$\lambda \simeq 0{\cdot}945 - \tfrac{1}{3}(d_{xt1} + d_{xb1})\frac{I_d}{I_{d1}} \quad . \quad . \quad (35)$$

For $p = 12$:

$$\cos \phi \simeq 0{\cdot}995 - \tfrac{2}{3}(d_{xt1} + d_{xb1})\frac{I_d}{I_{d1}} \quad . \quad . \quad (36)$$

$$\lambda \simeq 0{\cdot}990 - \tfrac{2}{3}(d_{xt1} + d_{xb1})\frac{I_d}{I_{d1}} \quad . \quad . \quad (37)$$

These four approximate formulae are shown dotted on the relevant Figures.

(6) CONCLUSIONS

It has been shown that the conventional theory of convertors can be extended so as to give information regarding the values of alternating voltage and power factor at convertor stations operating on a.c. systems having a finite short-circuit capacity. A formula for the additional direct-voltage regulation due to the impedance of the a.c. system has been derived. Finally, it has been shown that the power factors at convertor stations depend entirely on known plant and a.c. system data, so that acceptance-test measurements of these power factors are unnecessary.

(7) ACKNOWLEDGMENT

The author wishes to express his gratitude to Dr. J. C. Read, of the British Thomson-Houston Co., Ltd., for the many helpful suggestions received regarding the presentation of the paper. He is also indebted to Mr. C. T. Dillon, of ASEA, Ludvika, for his valuable assistance in connection with the preparation and translation of the paper.

NOTE ON "SYNTHESIS OF LADDER NETWORKS TO GIVE BUTTERWORTH OR CHEBYSHEV RESPONSE IN THE PASS BAND"*

The following note has been received from the author of the paper, Mr. E. Green.

Although the reasoning in this Monograph is conducted on a low-pass basis, the terminations are specified in terms of decrements, and not resistances. The former are very suitable for use with band-pass networks, since they are unaffected by changes in impedance level, but to use the solutions given for the low-pass prototype of Fig. 2(*b*), when the terminating resistances are unequal, we need a relation between their ratio and D, the ratio of the decrements. This can be obtained through expressions for the reflection coefficient, as outlined below.

Referring to the Monograph, for the low-pass network of Fig. 2(*b*), the first equation of Section 6.2, since $Z'_n = R_1$ at $x = 0$, can be written as

$$\rho_{x=0} = \frac{R_1 - R_n}{R_1 + R_n} = \left[\frac{\phi'_n(jx)}{\phi_n(jx)}\right]_{x=0} = \frac{U'_0}{U_0} \quad . \quad . \quad (A)$$

Type B Response.

Using eqns. (60*a*) and (60*b*),

$$\frac{R_1 - R_n}{R_1 + R_n} = \frac{(1 - D)^n}{(1 + D)^n}$$

so that

$$D = \frac{1 - \left(\dfrac{1 - R_n/R_1}{1 + R_n/R_1}\right)^{1/n}}{1 + \left(\dfrac{1 - R_n/R_1}{1 + R_n/R_1}\right)^{1/n}} \quad . \quad . \quad . \quad . \quad (B)$$

When dealing with a shunt input, G_n/G_1 will replace R_n/R_1 in eqn. (B). To avoid ambiguities as to sign, it is convenient to make R_n/R_1 (or G_n/G_1) and $D > 1{\cdot}0$ by suitable choice of input.

Having found D, design can begin, the process differing slightly according as n, the number of branches in the network, is to be odd or even.

When n is odd we can design the network with both terminations either series or shunt. If the choice is not obvious we can design both networks and then decide. When n is even the terminations are unlike and the higher value of resistance must always be at the shunt termination. Confining our attention to $D < 1{\cdot}0$, we get alternative designs by taking either end as input in turn, using R_n/R_1 for series input and G_n/G_1 for shunt input. Since the network numbering is reversed for the latter, both yield the same value of D.

Then if the bandwidth ω_β and tolerance V_p/V_β are specified, the decrements and coupling factors can be found from eqns. (23), and from these all the elements of the possible networks can be found.

* Monograph No. 88 R, January, 1954 (see 101, Part IV, p. 192).

Type C Response.

In a similar way when n is odd we find, by eqn. (A), for series terminations

$$\frac{R_1 - R_n}{R_1 + R_n} = \frac{U_0^{c'}}{U_0^c} = \frac{\sinh na'}{\sinh na} \quad . \quad . \quad . \quad . \quad (C)$$

Now $$\sinh na = \left(\left|\frac{V_p}{V_\beta}\right|^2 - 1\right)^{-1/2}$$

and is known when the tolerance V_p/V_β is known, so that

$$\sinh na' = \frac{1 - R_n/R_1}{1 + R_n/R_1} \sinh na$$

Thus a and a' can be found from tables.

When n is even,

$$\frac{R_1 - R_n}{R_1 + R_n} = \frac{\cosh na'}{\cosh na} \quad . \quad . \quad . \quad . \quad . \quad (D)$$

with

$$\cosh na = \left[1 - \left(\frac{V_\beta}{V_p}\right)^2\right]^{-1/2}$$

$$\cosh na' = \frac{1 - R_n/R_1}{1 + R_n/R_1} \cosh na$$

Thus again a and a' can be found.

Now, from eqns. (27*a*) and (43*a*) for all values of n,

$$\sinh a = d_1(1 + D) \sin\theta$$
$$\sinh a' = d_{11}(1 - D) \sin\theta$$

Therefore $$D = \frac{\sinh a - \sinh a'}{\sinh a + \sinh a'}$$

Having found D, the method of design follows that for type B response, except that we must use eqns. (27*a*) and (27*b*) to determine the decrements and coupling factors. With appropriate numbering the d's and K's will be the same for either series or shunt input, although the resulting networks will be different.

There is one further point to be noted: when n is even and the terminations are matched (i.e. $D = 1{\cdot}0$), then from eqn. (43*a*),

$$a' = 0, \quad \cosh na' = 1{\cdot}0$$

and from eqn. (D)

$$\frac{R_n}{R_1} = \frac{\cosh na - 1}{\cosh na + 1}$$

No greater values of R_n/R_1 with series input or G_n/G_1 with shunt input are possible.

ADDENDUM TO "THE PULSE TRANSFER FUNCTION AND ITS APPLICATION TO SAMPLING SERVO SYSTEMS"*

A method of determining the output of a sampling servo system between sampling instants was described very briefly in the Monograph. Although the treatment given is correct so far as it goes, it has since been found inadequate to cover cases in which the filter includes subsidiary feedback loops involving sampling. The imaginary delays must be inserted only in the main loop and not in any subsidiary loops.

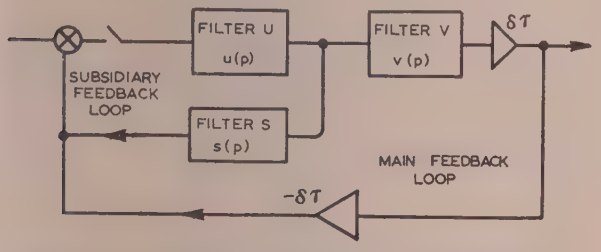


Fig. 1.—A stabilized servo system with imaginary delays.

Fig. 1 shows Fig. 3 of the Monograph redrawn to include a subsidiary stabilizing loop in which a modifying filter S may or may not be present. Such a system is best handled in two sections. The first, which we will call the controller, is the filter U as modified by the subsidiary feedback, and the second is the filter V. The signal input to V is a continuously ranging quantity, which means that it is necessary to use the "operational instruction" concept to specify the behaviour of the first section. The operational instruction is written in the form $\Theta(z) * \phi(p)$, in which $\phi(p)$ characterizes the law of variation of the controller output between sampling instants and $\Theta(z)/\Phi(z)$ is the pulse transfer function (p.t.f.) of the controller. The asterisk denotes that the sequence transform corresponding to $\phi(p)v(p)$ must be obtained before multiplication by $\Theta(z)$. The p.t.f. of all the equipment in the forward path is $W(z)$, given by

$$W(z) = \Theta(z) * \phi(p)v(p) \quad . \quad . \quad . \quad . \quad . \quad (1)$$

This does not yet include the effect of the imaginary delay $\delta\tau$, and so it relates only to the performance at the sampling instants.

Eqn. (26) enables the output between sampling instants to be evaluated, and for this purpose $W(z, m)$ is required. It will *not* normally be the column (iv) entry of the Table in the Appendix to the Monograph corresponding to $W(z)$, as found from eqn. (1), since the imaginary delay does not enter into the z-part of the operational instruction. Instead, the column (iv) entry corresponding to the product $\phi(p)v(p)$ is obtained and multiplied by $\Theta(z)$.

The following example illustrating the design of a zero velocity-lag servo mechanism will serve to clarify the method:

Filter V is a double integrator, $v(p) = Ap^{-2}$
Filter U is a clamp, $u(p) = (1 - \varepsilon^{-p\tau})p^{-1}$

Elementary considerations reveal the need for some stabilizing device, but the precise network function $s(p)$ has to be determined. To do this we proceed as follows:

Since the clamp output is constant between sampling instants, $\phi(p) = p^{-1}$ (see Table 1 of the Monograph). Hence

$$W(z) = \Theta(z) * Ap^{-3} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$= \Theta(z)\frac{\frac{1}{2}Az(z+1)}{(z-1)^3} \quad . \quad . \quad . \quad . \quad (3)$$

To have complete control over the behaviour of a third-order system a total of three constants capable of adjustment during design must be included in the system, i.e. $\Theta(z)$ must include two constants.

A further condition to be fulfilled is that the velocity lag must be zero. Hence the final expression for $W(z)$ must contain $(z-1)^2$ in the denominator, and therefore $\Theta(z)$ must contain $(z-1)$ in the numerator. The simplest expression which meets all these requirements and does not increase the order of eqn. (3) is

$$\Theta(z) = \frac{(z-1)(z-d)}{z(z+g)} \quad . \quad . \quad . \quad . \quad (4)$$

in which case

$$W(z) = \frac{\frac{1}{2}A(z-d)(z+1)}{(z+g)(z-1)^2} \quad . \quad . \quad . \quad (5)$$

The p.t.f. of the controller is therefore $\Theta(z)/\Phi(z) = (z-d)/(z+g)$. It is not difficult now to discover with the aid of the Table in the Appendix to the Monograph that this p.t.f. is realized if $s(p) = k/(p+\alpha)$, where $k = \alpha(g+d)$ and $d = \epsilon^{-\alpha\tau}$.

The characteristic equation of the system is $1 + W(z) = 0$, and by adjustment of the three available constants it can be matched term by term to any third-order equation known to provide suitable performance. The no-smoothing case, $z^3 = 0$, is the simplest and obtains when $A = \frac{5}{2}$, $g = \frac{3}{4}$ and $d = \frac{3}{5}$. The performance of the system as a whole at the sampling instants is given by the overall p.t.f., which is then

$$Y(z) = \tfrac{5}{4}z^{-1} + \tfrac{1}{2}z^{-2} - \tfrac{3}{4}z^{-3} \quad . \quad . \quad . \quad . \quad (6)$$

To find the output between sampling instants we require $W(z, m)$, which is obtained from eqn. (2) by using the column (iv) entry corresponding to p^{-3}.

i.e. $W(z, m)$

$$= \frac{\frac{1}{2}A(z-d)}{z(z+g)(z-1)^2}[m^2z^2 + (1 + 2m - 2m^2)z + (1-m)^2] \quad (7)$$

The same values as previously must, of course, be used for the design constants, and the overall performance is then given by

$$Y(z, m) = \tfrac{1}{4}[5m^2z^{-1} + (5 + 10m - 13m^2)z^{-2} + (2 - 16m + 11m^2)z^{-3} - 3(1-m)^2z^{-4}] \quad . \quad (8)$$

As m varies from zero to unity the coefficient of the first term $(5/4m^2)$ shows how the response to a unit sample input varies

* Barker, R. H.: Monograph No. 43 M, July, 1952 (see 99, Part IV, p. 302). The addendum was first received 10th June, and in revised form 9th July, 1955.

throughout the first sampling interval. The coefficient of the second terms shows similarly the response throughout the second sampling interval, and so on. To obtain the response to any other input function, $Y(z, m)$ is simply multiplied by the sequence transform of that function.

In this example there has been no cancellation of poles and zeros. If, however, the constant g were given the value unity in eqn. (5) it would appear superficially that a simpler system of second order instead of third would result. The term $(z + 1)$ in the numerator only holds for the special values of m equal to zero or unity, and therefore such cancellation is not a permissible procedure. It would, in fact, serve only to conceal oscillations of the type mentioned in Section 4.3 of the Monograph. Cancellation of poles and zeros which are inside the unit circle will conceal a decaying oscillation. Cancellation of poles and zeros is, in general, permissible only if it holds good for all values of m.

Acknowledgments are due to the Controller, H.M. Stationery Office, for permission to publish this note and to Dr. J. H. Westcott who first pointed out the inadequacy of the Monograph.

INDEX TO VOLUME 102, PART C

1955

ABBREVIATIONS

(P)—Paper.

(D)—Discussion.

Printed in Great Britain by
UNWIN BROTHERS LIMITED, WOKING AND LONDON

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

ISSUED IN THREE PARTS AS FOLLOWS:

Part A. POWER ENGINEERING (*February, April, etc.*)

Part B. RADIO AND ELECTRONIC ENGINEERING (*January, March, etc.*)

Part C. INSTITUTION MONOGRAPHS (*March and September only*)

PART C—MONOGRAPHS

SEPTEMBER 1955

CONTENTS OF THIS ISSUE

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part.
Example.—SMITH, J.: "Reflections from the Ionosphere," *Proceedings I.E.E.*, Paper No. 3001 R, December, 1954 (102 B, p. 1234).

Published by The Institution, Savoy Place, London, W.C.2. *Telephone:* Temple Bar 7676. *Telegrams:* "Voltampere, Phone, London."
Printed by Unwin Brothers Limited, Woking and London.